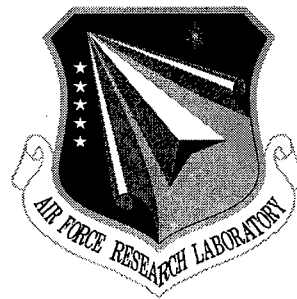


AFRL-IF-RS-TR-1999-203
Final Technical Report
September 1999



MULTISTAGE OPTICAL INTERCONNECTS FOR PARALLEL ACCESS OPTICAL MEMORY

University of California

Y. Fainman
Paul E. Shames

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK**

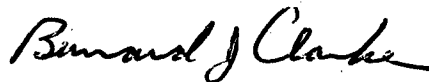
DTIC QUALITY INSPECTED 4

19991222 072

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-1999-203 has been reviewed and is approved for publication.

APPROVED:



Bernard J. Clarke
Project Engineer

FOR THE DIRECTOR:



John V. McNamara
Technical Advisor
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFED, 32 Brooks Road, Rome, NY 13441-4114. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 1999		3. REPORT TYPE AND DATES COVERED April 1995 - January 1998
4. TITLE AND SUBTITLE MULTISTAGE OPTICAL INTERCONNECTS FOR PARALLEL ACCESS OPTICAL MEMORY			5. FUNDING NUMBERS C - F30602-95-C-0065 PE - 62702F PR - 4594 TA - 15 WU - P1	
6. AUTHOR(S) Y. Fainman, Paul E. Shames				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California San Diego Department of Electrical and Computer Engineering 9500 Gilman Drive La Jolla CA 92093-0407			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/IFED 32 Brooks Road, Rm 321A Rome NY 1344-4114			10. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-IF-RS-TR-1999-203	
11. SUPPLEMENTARY NOTES AFRL Project Engineer: Bernard J. Clarke/IFED/(315) 330-2106				
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This contract demonstrated an optical multistage interconnection network (MIN) for interfacing and distribution of parallel access optical memories. This demonstration consisted of a folded, free-space optical MIN using a novel folded dilated bypass-exchange switch (DBS) built using birefringent computer generated holograms (BCGH) and polarization rotator elements. By utilizing the three-dimensional functionality of the optical elements, the DBS elements can be stacked in the vertical dimension by folding the switch along its central line of symmetry. The interconnection between multiple DBS is also folded, forming a compact, transparent optical MIN network.				
14. SUBJECT TERMS Multistage Interconnection Network (MIN) Birefringent Computer Generated Hologram (BCGH)			15. NUMBER OF PAGES 140	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Abstract

We have completed our research on a transparent optical multistage interconnection network (MIN) for interfacing and distribution of parallel access optical memories. This network allows for the transparent transmission of optical data, which is required for many applications such as high speed image computing, data base search, digital library, and telemedicine. The transparent switching fabric of this network is based upon the unique technology of birefringent computer generated holograms (BCGH). BCGH technology, network design and initial system architecture studies were first developed under contract #F30602-91-C-0094, where we showed the viability and capabilities of BCGH based switching elements.

During this contract we have designed and implemented a folded free-space optical MIN using a novel folded dilated bypass-exchange switch (DBS) built using BCGH and polarization rotator elements. The DBS allow for the elimination of first-order cross-talk due to inaccuracies of polarization rotation and diffractive element fabrication errors. By utilizing the three dimensional functionality of the optical elements, the DBS elements can be stacked in the vertical dimension by folding the switch along its central line of symmetry. The interconnection between multiple DBS is also folded, forming a compact transparent optical MIN package. This compact optical system design is easily aligned and the use of space-variant lenslets permits implementation of arbitrary network architectures and interconnection patterns. Additionally, the use of patterned micro-mirrors to fold the DBS allows for spatial filtering of the undesired high diffraction orders, thereby decreasing the cross-talk.

Fabrication and characterization of the folded transparent optical MIN demonstration system has highlighted the advantages of this design. Using BCGH elements with measured signal-to-noise ratios (SNR) of 30:1, the fabricated folded 2×2 DBS improved SNR to 60:1. Increasing the scaling to a 4×4 MIN resulted in SNR of 120:1. Therefore, using the folded MIN design, system scaling is limited only by insertion losses, which may be mitigated with the use of optical amplifiers. To handle the path contentions that may arise in these types of networks we have developed a novel switching protocol named the gated-hold protocol. A stochastic model has also been developed that allows for the rigorous performance analysis of these algorithms.

We have also continued developing BCGH technology by using a highly birefringent substrate material allowing fabrication of single substrate polarization selective elements designed with a novel multiple order delay (MOD) approach. Additionally, form birefringent computer generated holograms (FBCGH) have been developed. The FBCGH use subwavelength gratings to generate birefringence several times greater than possible using natural anisotropic materials, demonstrating SNR greater than 250:1. To meet the switching speed requirements of the next generation MIN systems, we have also conducted extensive modeling and testing of PLZT based phase modulation devices for the implementation of high-speed polarization rotation devices. Using readily available MOSFET drivers, we have developed PLZT based devices working at 10 MHz, i.e. three orders of magnitude faster reconfiguration time of optical MIN compared to a system using current ferroelectric liquid crystal technology.

Table of Contents

Abstract	i
Table of Contents	ii
Table of Figures.....	iii
1. <u>Overview</u>	1
2. <u>System Design of the Interconnection Network</u>	3
3. <u>Switching Protocols</u>	9
4. <u>Polarization Selective Computer Generated Holograms</u>	11
5. <u>Polarization Rotation Devices</u>	14
6. <u>System Demonstration</u>	18
6. 1. Folded Multistage MIN.....	18
6. 2. Reduced Cross-talk MIN.....	21
7. <u>Conclusions</u>	26
8. <u>References</u>	28
<u>Appendices</u>	30
A. 1. Polarization-controlled multistage switch based on polarization-selective computer-generated holograms.....	30
A. 2. Dynamic Performance of a Circuit-Switched Interconnection Network	45
A. 3. Request Resubmission in a Blocking, Circuit-Switched, Interconnection Network.....	52
A. 4. Single Substrate Birefringent Computer Generated Holograms	65
A. 5. Multiple Order Delay Holograms for Polarization and Color Selectivity.....	69
A. 6. Form-birefringent computer-generated holograms	74
A. 7. Design, fabrication, and characterization of form-birefringent multilayer polarizing beam splitter.....	78
A. 8. Fabrication, modeling, and characterization of form-birefringent nanostructures	89
A. 9. Polarizing Beam Splitters Constructed of Form-Birefringent Multilayer Gratings.....	93
A. 10. Modeling and Optimization of Electro-Optic Phase Modulator	102
A. 11. Modeling Electric Field Induced Effects in PLZT EO Devices.....	113
A. 12. Schematic for push-pull MOSFET driver	117
A. 13. System design and component layout for optical MIN demonstration	119
A. 14. Folded Free-Space Polarization-Controlled Multistage Interconnection Network.....	121

Table of Figures

Figure 1. Transparent Optical Multistage Interconnection Network	2
Figure 2. Bypass-exchange switch based using two BCGH	4
Figure 3. Optical implementation of DBS switch is made up of four BCGH elements	6
Figure 4. Folded optical dilated bypass exchange switch	7
Figure 5. Folded optical multistage interconnection network.....	8
Figure 6. Micro-graph of FBCGH linear grating.	13
Figure 7. Bottom curve shows the 1 MHz driving and top curve shows the V_{π} response of the PLZT	16
Figure 8. EO response of PLZT 8.8/65/35	17
Figure 9. (a) Transmission intensity through crossed polarizers. (b) Transmission intensity when neighboring elements are in 'off' state.....	17
Figure 10. Folded DBS output	20
Figure 11. Microscopic photograph of new single BCGH element.....	21
Figure 12. (a) Output of new combined BCGH element. (b) Spatial filtering using micro-mirrors	22
Figure 13. Picture of second generation folded MIN experimental system.....	23
Figure 14. Output from DBS fabricated using new BCGH elements	24
Figure 15. Output packets from 4×4 switch with two input signals	25
Figure 16. Switching a single DC input between eight output channels.....	25

1. Overview

This final report outlines the past two-year's research on development of a compact folded optical multistage interconnection network (MIN) for parallel access and distributed optical memory. For convenience, publications [1-12] and schematics are included as appendices to this report.

Electronic multiplexing can be used to access distributed memory devices (optical and magnetic disk arrays, multi-head disks, etc.), providing data rates at 1 to 10 GHz. Existing memories are commonly arranged in a distributed environment, where the use of optical fiber transmission lines becomes the most effective medium. The interfaces to such memory systems would benefit from transparent optical switching fabric that allow the transmission of information, in any data format, at rates limited only by the speed characteristics of the electronic devices at the transmitter (e.g. memory system node) and the receiver (e.g. user node).

The recent advances in optical amplifiers have increased interest in such transparent optical networks. Additionally, since polarization compensation in single mode fiber [13] allows automatic and stable control of the polarization states of transmitted optical signals, it may enable utilization of polarization dependent all-optical switching fabric. Such polarization switching has been proposed for 'free-space' MIN for switching and multiprocessor interconnections [14-16]. Leveraging off of birefringent computer generated hologram (BCGH) research done under Contract #RADCF-30602-91-c-0094, we have recently demonstrated a 4x4 optical MIN (see Appendix 1). Interconnection controls, which necessarily operate at much slower rates, can be executed electronically. Such a transparent network can accommodate bandwidths several orders of magnitude greater than possible with electronic systems. This transmission bandwidth capability is comparable with parallel access optical memories that are expected to provide aggregate bandwidths approaching Tbits/s.

A block diagram of such a memory distribution system is shown schematically in Figure 1. The control of the switching fabric and its input and output nodes can be performed using existing electronic network technology (e.g. a transparent optical MIN may work in parallel to existing electronic MIN, where the electronic network handles the transmission of low priority

data and control of the high-speed optical network). The transparent optical switching fabric is implemented with polarization selective diffractive optics and polarization rotator arrays being developed at UCSD. The primary goal of this project is to use these technologies to implement a compact and scalable high-speed interconnection system for optical memory access.

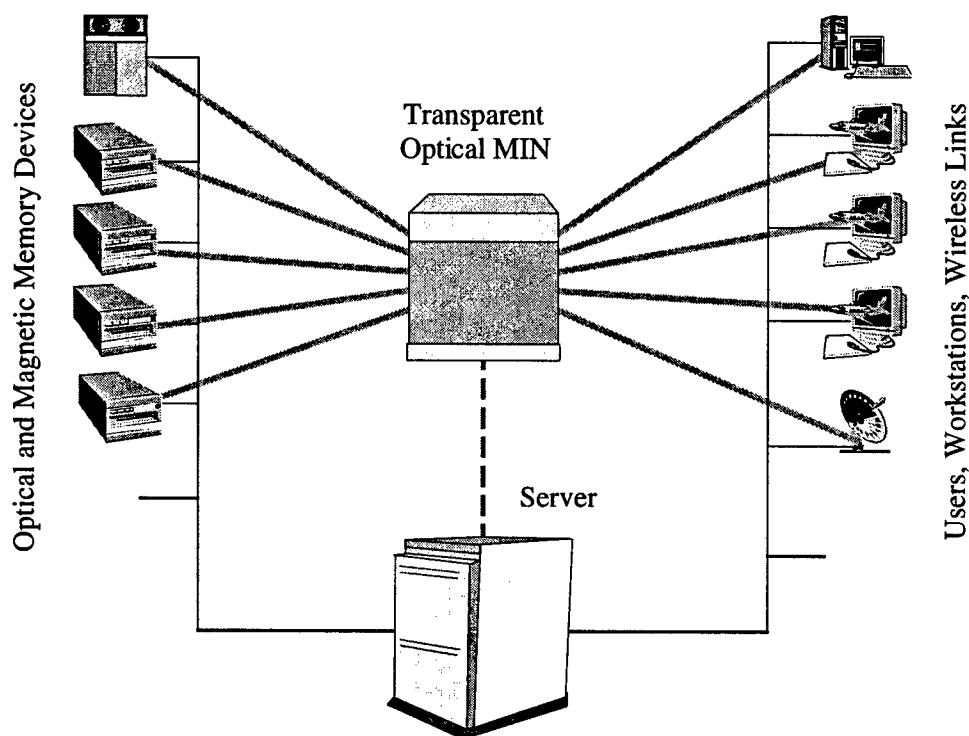


Figure 1. Transparent Optical Multistage Interconnection Network can overlay and is controlled by an existing electronic network made up of various memory device and user nodes and a server.

During the course of this contract we have focused our research on the creation of a demonstration compact optical MIN as well as the development of the novel required for its operation. In particular we have made advances in four areas: optical MIN system design, switching network protocol, and developing the technologies of polarization selective diffractive optical elements (DOE) and high-speed polarization rotation device arrays.

One of the novel systems that we have developed is the folded dilated bypass-exchange switching (DBS) fabric. This unique folded design allows us to take advantage of the symmetries of the optical switching system. The folded architecture places all like elements into single 2D arrays, thus reducing the system space requirements as well as simplifying system alignment. The

system demonstration has been fully designed, fabricated and characterized. The demonstration system utilizes our current accomplishments in developing the underlying technologies. In further developing BCGH we have made significant improvements in the concept developed under contract #RADCF-30602-91-c-0094. We have designed, fabricated and characterized a single substrate BCGH that resolves many of the fabrication issues that were seen in two substrate BCGH. We have also developed rigorous design tools to investigate design and fabrication tolerances for polarization selective DOE. Furthermore, we have designed and fabricated single substrate form birefringent polarization selective DOE. Finally, we are modeling, fabricating and characterizing a high-speed array of polarization rotation devices that will provide several orders of magnitude improvement in speed over current polarization rotation technology.

In this report we will summarize our progress on each of the four major areas of this research that were described above and also present results of the system demonstration: The following section will describe the basic concepts and development that has gone into the system design. In Section 3 we describe the switching protocol strategies that have been developed for this project. Section 4 will discuss the work done on polarization selective computer generated holograms. In Section 5 we describe the continued work on polarization rotation devices. Section 6 describes the fabrication and characterization of the two system demonstrations. In Section 7 we conclude this report with an assessment of the work that has been performed under this contract.

2. System Design of the Interconnection Network

The interconnection network module that has been designed and constructed enables dynamic switching between input and output optical signals. The system is transparent to the optical signals, i.e. the signals are not converted to the electrical domain for digital switching and are not remodulated on an optical carrier. Due to this transparent feature, the system is independent of the bit-rate, or bandwidth, of the signals. The switching is performed by an applied electrical signal, coming from an auxiliary module, running the protocol and routing algorithms (see Section 3).

The interconnection network module is based on several enabling technologies for controlling the phase front and polarization of optical beams. The first is computer generated holograms (CGH), which provide general optical element functionality independent of the polarization of the light. The second enabling technology which we are taking advantage of is polarization selective computer generated holograms. Similar in concept to the CGH, two general functionalities can be encoded onto it, one for each polarization state. Under contract #RADCF-30602-91-c-0094 we succeeded in designing, fabricating and experimentally evaluating these elements by constructing special computer generated holograms in birefringent material (BCGH). The third technology is polarization modulators, that can switch between polarization states, by an applied electric field. By integrating these technologies we have formed the basis for a transparent optical switch, whose precise functionality is dictated by the choice and placement of these elements.

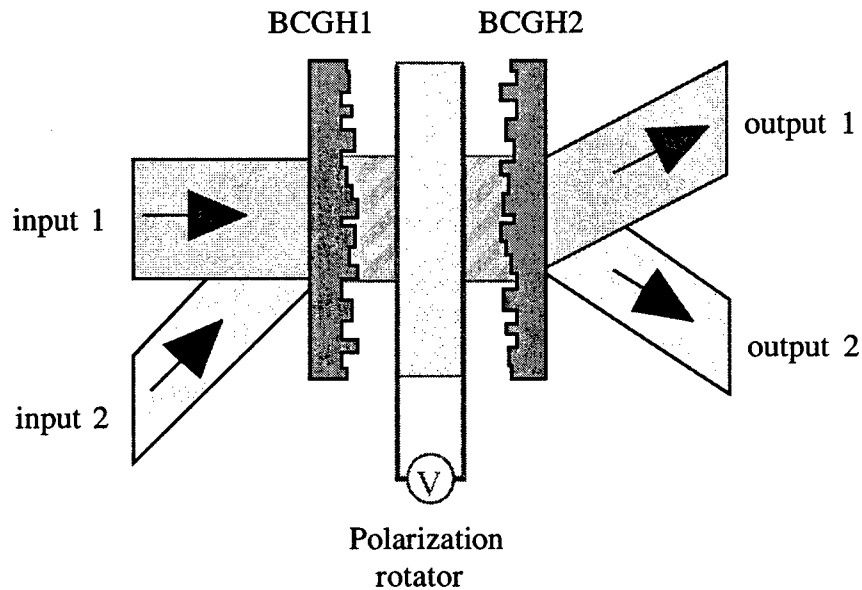


Figure 2. Bypass-exchange switch based using two BCGH and one polarization rotation element. Input beams are orthogonally polarized.

In the past we have demonstrated the bypass-exchange switch, a fundamental switch for two signals [17]. Two BCGH elements and an electrooptic (EO) polarization rotator can be used to construct a 2×2 optical BES (see Figure 2). The first BCGH element combines and focuses two inputs into the polarization rotator, which either exchanges their polarizations or not. The second

BCGH separates and directs the outputs to different destinations. Inaccuracies of the polarization rotator can result in cross-talk in this implementation of the BES. The polarization rotator can be characterized by an associated error of δ in the rotation angle, which results in a cross-talk term proportional to $\sin(|\delta|)$. The BCGH elements can be described by an associated cross-talk, ϵ , due to fabrication errors such as etch depth and mask alignment. The combined cross-talk component at the output of the BES is proportional to $|\delta| + |\epsilon|$, assuming $\delta, \epsilon \ll 1$. The signal-to-noise ratio (SNR) of a MIN can be described by

$$\text{SNR} = \log_{10}\left(\frac{1}{\delta_c}\right) - \log_{10} S \quad (1)$$

where, $\delta_c = |\delta| + |\epsilon|$ and S is the number of interconnection stages [1]. For scalability of MIN network size (i.e. S is growing) the cross-talk, δ_c , of each stage must be reduced to achieve necessary SNR.

The dilated bypass-exchange switch (DBS), which utilizes a more complex structure, performs the functionality of the BES with improved cross-talk performance [18]. The DBS, which has two input and two output signals, is comprised of four 1×2 elements coupled together. The structure of the DBS guarantees that each bypass exchange switch has only one signal propagating through it, and that the majority of the cross-talk terms exit from the ports that are not utilized. It can be shown that the remaining cross-talk is now proportional $\delta^2 + \epsilon^2$. Under the assumption $\delta, \epsilon \ll 1$, cross-talk is greatly decreased and SNR increased.

A free-space DBS can be implemented with a combination of lenslet, BCGH and polarization rotator elements (see Figure 3). As opposed to the bypass-exchange switch implementation, the input signals to the DBS are spatially displaced from each other and may be directed using off-axis Fresnel lenslets fabricated from isotropic material. This allows the single-substrate BCGH, which can be highly sensitive to fabrication error, to be made using a relatively simple linear phase encoding. One polarization rotator, whose polarization state determines whether the switch will function in bypass or exchange mode, can control both signals' polarization. The first BCGH elements have one phase encoding for a vertical polarization state (e.g. bypass mode) and a different phase encoding for the other orthogonal polarization state. The second set of BCGH

elements have the conjugate functionality as the first set, serving to recollimate the beams. The second polarization modulator is set in the same state as the first, so that the output polarization state is also identical to the input state. The final lenslets direct the output beams according to the interconnection pattern required by the system architecture. Linear cross-talk terms, which exit the DBS with a polarization state that is orthogonal to that of the desired signals, can be filtered out with a polarizer.

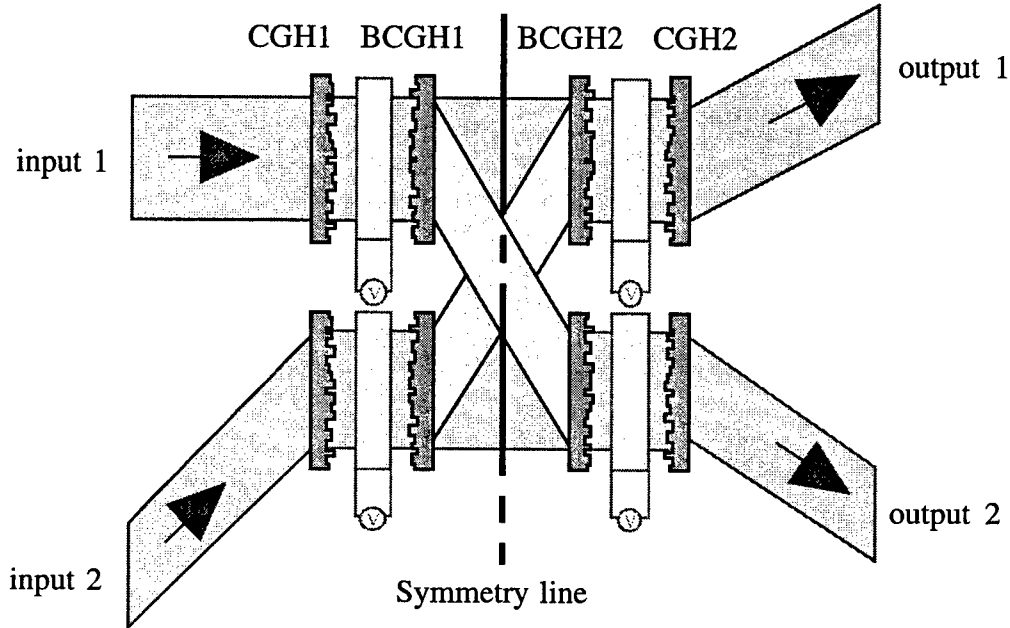


Figure 3. Optical implementation of DBS switch is made up of four BCGH elements. The order of elements on the left and right hand sides are symmetric around the central line of symmetry. Input beams are independent.

The DBS complexity, while mitigating the linear cross-talk problem, increases the number of components required for the same functionality as the BES. However, taking advantage of the symmetry of the DBS (see Figure 3) and the three-dimensional functionality of our free-space optical elements can reduce the complexity of these switches. This is done by introducing a propagation direction component along the vertical axis, i.e. a small incidence angle, as well as placing a mirror at the line of symmetry. The input beams will pass through a lenslet-rotator-BCGH combination at one elevation and react according to the encoded information at that location. Upon reflection from the mirror the beam passes through another BCGH-rotator-lenslet combination at a lower elevation (see Figure 4). By folding the switch in this manner, similar

elements (i.e. BCGH, lenslets and polarization rotators) are located in the same plane. Therefore, a DBS may be fabricated using a mirror and 2×2 arrays of BCGH, lenslets and polarization rotator elements.

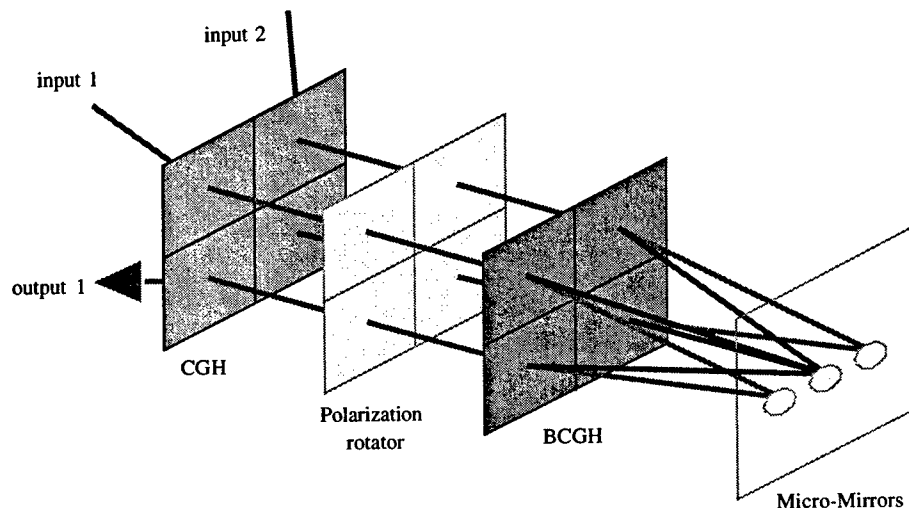


Figure 4. Folded optical dilated bypass exchange switch locates like elements into 2D arrays. Micro-mirrors can be used to filter high diffraction order noise.

The advantage of this folding technique is further enhanced when applied to an optical MIN. By placing a mirror at the output of the first folded DBS, the beam will reflect back at a lower elevation and be coupled into subsequent DBS located below the first. In this manner all similar elements of multiple DBS may be combined into two-dimensional arrays, minimizing the number of components required for the entire MIN: a single lenslet array, a BCGH array, a polarization rotator array and a pair of folding micro-mirror arrays. A folded optical MIN can be packaged as a resonator, where each round trip represents a stage, and all stages are stacked vertically (see Figure 5). An input signal beam enters the system at a small angle and reflects through a prescribed number of stages before exiting in the desired spatial output channel.

For compactness and optimal use of available fabrication technologies, we use computer-generated hologram (CGH) off-axis Fresnel lenslets. BCGH and CGH lenslets are diffractive elements, whose diffraction efficiency dictates the amount of unwanted higher diffraction terms produced. Using continuous mirrors, these unwanted orders may propagate within the MIN, resulting in additional cross-talk. However, by using micro-mirrors deposited on a transparent

substrate, only the desired diffraction terms from the BCGH will reflect back for further propagation, while the unwanted noise terms exit the system.

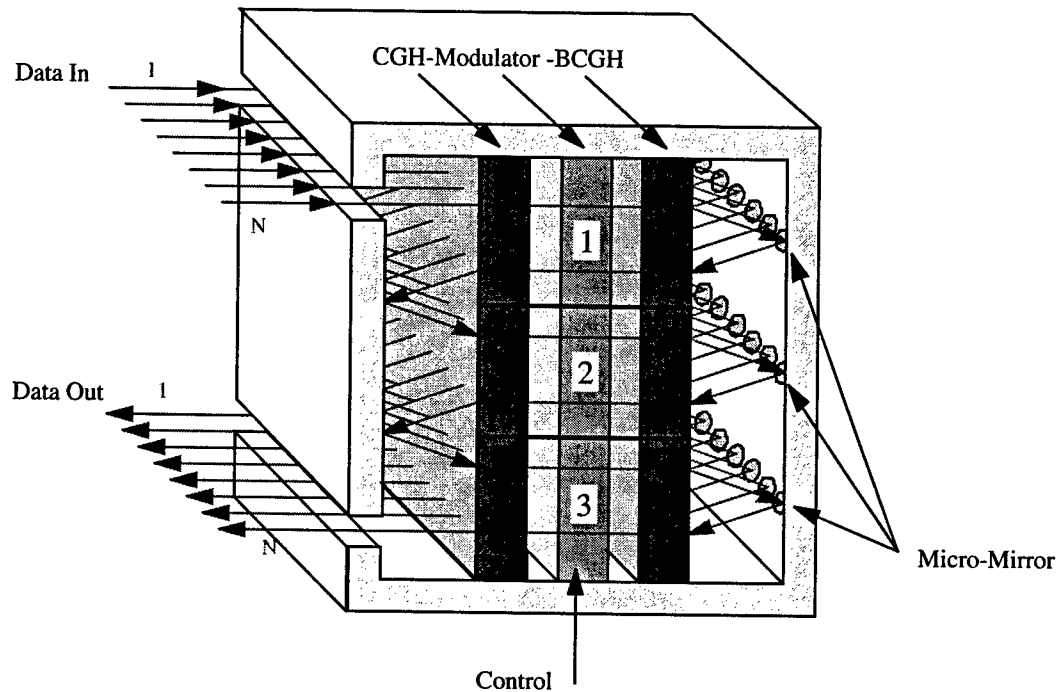


Figure 5. Folded optical multistage interconnection network shows compact packaging using 2D arrays of optical elements. For an 8×8 folded MIN there are three layers of DBS, which are shown as separate rows on the 2D arrays.

The arrangement of the optical elements in 2-D arrays also allows for relatively simple alignment of the system components. Correct alignment will dictate that during each pass through the cavity the beams will land on the correct elements. Based on geometrical ray tracing, the displacement of each beam from its correct position and the size of the beam at the BCGH elements (i.e. larger or smaller than the predicted size at the element) will indicate which optical elements (BCGH, micro-mirrors, etc.) are incorrectly positioned. Since the mirror planes are mostly transparent, beam propagation within the cavity can be viewed with the use of external imaging optics and a CCD camera. The beam size and position was monitored in situ allowing for accurate alignment of optical elements and mirror planes.

The use of space-variant lenslets in each polarization selective element allows for the design of arbitrary connection patterns such that any network topology may be implemented. In the

folded optical MIN, the number of channels and the interconnection architecture used dictates the size of the arrays but does not increase the number of components. For example, an 8×8 optical MIN architecture (of $\log_2 8 = 3$ stages) requires arrays of size 8×6 in BCGH and polarization rotator elements.

3. Switching Protocols

A fully connected network provides the connectivity such that any input can connect to any output. These networks can be blocking if there are internal contentions for links with existing network connections. Networks are rearrangeably non-blocking if any idle input may be connected to any idle output provided that we may rearrange existing connections [1]. A rearrangeably non-blocking network can also be achieved by cascading two fully connected networks. Thus, fully connected networks are smaller and have simple routing algorithms, but may incur blocking situations which need to be resolved on the protocol level. Rearrangeably non-blocking networks can accommodate any interconnection pattern at the expense of a larger network. Routing is more complex for these networks, and short breaks in communication may result when reconfiguring a network.

For applications that require a MIN system of limited size and complexity it may be convenient to implement a fully connected network. For centralized control of this blocking system we have developed a novel switching protocol to handle routing path contentions. The basic blocking system architecture is that there are N users accessing M data sources (e.g. optical or magnetic disks) via a server regulated network. The problem is that given n arrivals of information at times t_1, t_2, \dots, t_n at any node in an interconnection network the protocol controlling the network must decide the order of service of all N users in order to avoid contention and provide quality of transmission. Previous work on these kinds of blocking or conflicting transmission problems have used non-real time and real time scheduling disk access policies. The negotiated techniques, to assure stability of the network, only give conditions on buffer size and load limited to a single disk without describing the transient behavior of the system. The analysis of these policies has generally been done by deterministic approaches which doesn't allow a rigorous testing of the suitability of these protocols. Moreover, no tangible results

have been found in the analysis of these protocols in scheduling and buffer allocation for disk arrays.

The gated-hold protocol that has been developed by Paul Dietrich and Prof. Ramesh Rao (see Appendix 2 and 3) has been rigorously analyzed using stochastic methods. This protocol, in a cycle, will guarantee the delivery of all requests. The gated-hold protocol compares favorably to other blocking network protocols. However, the overhead imposed by the gated-hold protocol is characterized by the hardware complexity. Depending on the hardware configuration a different method of analysis may be required.

Most studies that have been done in scheduling of disk arrays have been done using computer simulations or deterministic approaches, which doesn't always describe the random nature of communication systems, which are characterized by burstiness at peak loads. We are developing a stochastic model that will give us a better representation of these systems. This will allow us to quantify the characteristics of a given system and then to derive an optimal protocol.

Another issue that we continue to investigate is the seek-delay time of a disk. Data is stored in blocks on the disk and a certain time is needed for the disk head to locate that block on the disk. We are investigating disk access policies on many issues in relation to minimizing the amount of overall seek delay. These issues include storing information contiguously or non-contiguously, coordination of disk movement and service of requests on single and multi-disk systems. As an example, with the assumption of uniform arrivals of N users (which is considered the worst case) accessing a single disk, we have established the expected dead-time expression for a range of scheduling policies. Given the dead-time that we can tolerate and the rotation speed of the read and write disk head, we can find the buffer size requirement (or vice versa).

We are also interested in multimedia networks, which may impose a totally new set of requirements onto a system protocol. The combination of text, image, video and voice may require varying information packet sizes, transmission and access speeds as well as varying priorities for different information types and packets.

4. Polarization Selective Computer Generated Holograms

One of the key technological components in our transparent photonic MIN is the birefringent computer generated hologram (BCGH) array. These diffractive optical elements (DOE) have independent impulse responses for two orthogonal linear polarizations. Two degrees of freedom are required in order to encode two different phase functions into one diffractive element. In its original design, the BCGH consists of two surface relief substrates with at least one of them birefringent. The first birefringent substrate introduces a relative phase delay in each pixel between the two orthogonal linear polarizations. The second isotropic substrate controls the absolute phase of both polarization components. The use of two independent etch depths provides the two degrees of freedom necessary for independent impulse responses. Previously, we have demonstrated BCGH with different functionality using two birefringent LiNbO_3 substrates [17]. The fabricated BCGH elements showed high diffraction efficiency (as high as 60%) and large polarization contrast ratio ($> 100:1$).

Due to the special configuration of a two substrate BCGH element, different analyses are required to understand the relationship between the performance of a BCGH and the imperfection introduced during the fabrication processes. Using both scalar diffraction analysis (i.e., Fourier analysis) and rigorous vector field analysis (i.e., rigorous coupled wave analysis), we found that two substrate BCGH elements are more sensitive to fabrication imperfections compared to regular diffractive optical elements. Our numerical analyses indicate that to construct a high performance two substrate BCGH, tighter fabrication tolerance is required in terms of exposure dosage, etch depth accuracy and alignment accuracy.

The difficulties we encountered in making two substrate BCGH serve as one of the motivations for us to investigate new approaches to make a BCGH element. One approach we took is to increase the etch depth in a single birefringent substrate (see Appendix 4) so that the phase delay caused by the etched pixel compared to an unetched one is more than 2π . We name such elements multiple order delay (MOD) holograms. Inside a MOD single substrate BCGH element, each pixel of the microstructure is deep-etched such that propagating optical waves will exhibit multiple periods of phase delays. Therefore, the variable order of phase delay provides us

another degree of freedom, in addition to the etch depth, to encode two phase functions. Compared to a two substrate BCGH, a MOD element has the advantages of simpler element configuration, relative ease of fabrication, and may yield higher performance in terms of polarization contrast ratio and diffraction efficiencies.

We designed, fabricated and evaluated both experimentally and numerically a binary phase level MOD with polarization selectivity. The substrate material was YVO_4 which has refractive indices of $n_o = 2.0241$ and $n_e = 2.2600$ at a wavelength of $0.5145\mu\text{m}$. The hologram is designed to be a polarization beam splitter, i.e., transmit one polarization straight and deflect the orthogonal polarization at an angle. The element was etched to a designed depth of $1.032\mu\text{m}$ using ion beam etching. The diffraction efficiencies measured under ordinary and extraordinary polarization illumination are 70.8% into the zero order, 37.4% into the +1st order, and 38.9% into the -1st order. The measured polarization contrast ratios are 79.7:1 at zero order, 33.0:1 at +1st order and 32.5:1 at -1st order. We also simulated the performance of a MOD using rigorous coupled wave analysis (RCWA). The simulation results agreed with our experimental evaluation well.

In the MIN demonstration we use four phase level MOD elements. The diffraction efficiency was experimentally measured at 55.3% for the 0 order and 45.3% for the +1 order, with extinction ratios of 10:1 and 30:1, respectively. The diffraction efficiencies into the +1 order were better than the binary phase element, but significantly below the theoretically predicted efficiency of 80.5% diffraction efficiency for a four phase level DOE. The relatively poor efficiency can be attributed to errors in etch depth during the fabrication process. In the MOD approach, etch depth must be controlled with a higher degree of accuracy than that used for conventional DOE. This is because the error introduced by the over/under etch is determined by the ratio of the percent of etch error to a fraction of the total etch depth that is, in effect, responsible for encoding the desired phase values on a given pixel. Therefore, the multiple phase level elements were more susceptible to fabrication errors due to an inconsistent ion beam etching process.

We have also applied the multiple order delay approach to single substrate BCGH which are

color selective (see Appendix 5). We designed and fabricated a binary phase level color selective beam splitter for wavelengths 1.30 μm and 1.55 μm . The substrate material used was BK7 glass which has index of refraction 1.5027 at 1.30 μm and 1.5004 at 1.55 μm . The diffraction efficiency was 39% into the +1 and -1 diffraction orders and 0.83% 0 order transmission for 1.55 μm wavelength. For 1.30 μm wavelength light the 0 order transmission was 83% with less than 1.2% diffraction into the other orders.

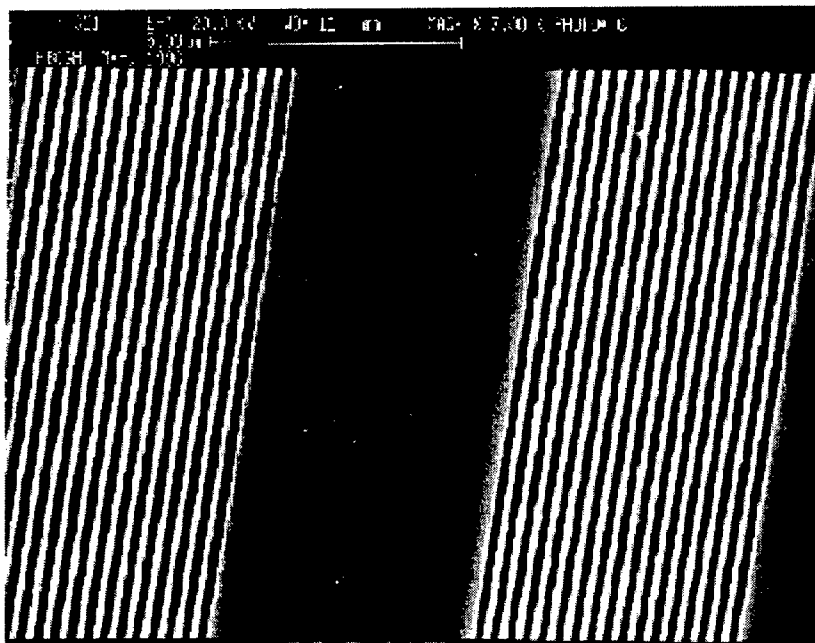


Figure 6. Micro-graph of FBCGH linear grating. Phase encoding is performed using a form-birefringent based sub-wavelength grating structure.

Another approach to BCGH technology can be implemented using form birefringent subwavelength gratings (FBCGH) on single isotropic substrates (see Appendix 6, 7, 8 and 9). Electric fields parallel to the grating (TE polarization) and perpendicular to the grating (TM polarization) need to satisfy different boundary conditions, resulting in different effective refractive indices for the two orthogonal polarizations. It has been shown that the birefringence possible using this approach can be several times greater than naturally occurring materials. Using a combination of effective medium theory (EMT) and RCWA we designed, fabricated and characterized a linear grating using a binary phase FBCGH (see Figure 6). The experimentally measured diffraction efficiencies were 75% for the 0 order, 41.4% for the +1 order and 44.2% for

the -1 order, with extinction ratios of 88:1, 275:1 and 99:1, respectively. The form birefringent structures also serves as an antireflection coating, explaining the slightly higher measured diffraction efficiencies compared with that predicted by scalar diffraction theory for a binary phase element (40.5%).

5. Polarization Rotation Devices

The reconfiguration of our transparent switching fabric is accomplished with electrooptic (EO) polarization rotation devices. The 'folded' configuration of our network demonstration requires a 2D array of elements, which can currently be best implemented using a commercial liquid crystal (LC) device. However, our research continues on development of a new generation of polarization rotation devices implemented using electrooptic crystals or ceramics, such as lanthanum-modified lead zirconate titanate (PLZT).

The implementation of our transparent switching fabric for use in an optical MIN can take various forms. As an example, a set of users may wish to access optical information stored on a set of optical disks. The MIN can be configured, allowing for routing conflicts (i.e. blocking), so that information can pass from the disks to the users as fast as the signal can be generated. The limit to the speed of access comes from the active components of the system. At present the greatest latency in such a system would come from the access time of the optical disks (on the order of milliseconds). However, in the future other optical memory devices may be available which will allow much faster access times. In that case the switching speed of the polarization rotation devices within the MIN could be the limiting factor.

The ferroelectric LC device from Display Tech that we are currently using has a response time on the order of a hundred microseconds. The other device parameters that are important are the contrast ratio, transmittance and cross-talk between array pixels. In testing of a 10×10 array designed for infrared operation we found that the cross-talk between 1 mm^2 pixels was minimal and the contrast ratio was as high as 100:1. Preliminary testing of a 10×10 array designed for visible light has shown similar characteristics. Response time (as well as contrast ratio) of the material is a function of the speed and voltage at which the device is run. Operating at 67 Hz, with ± 15 Volts applied, the infrared device had a response time of 135 μs .

In order to decrease the reconfiguration time of the polarization rotation portion of the MIN, alternative materials will need to be used. We have been investigating the feasibility of using PLZT, an electrooptic ceramic material, as a replacement for the ferroelectric liquid crystal. PLZT 9.0/65/35 has a very strong electrooptic effect combined with relatively slim-loop hysteresis, which can be used to achieve polarization rotation at reasonable switching voltage ($V_{\pi} \approx 150$ Volts) and has a reported response time on the order of 100 nanoseconds.

Our investigation has found that the design of an efficient polarization rotation device based on PLZT must contend with several critical design parameters. For thin wafers, the transmittance of the ceramic is very high, almost 100%, however using thin wafers entails using surface electrodes which result in curved electric fields and inhomogeneous phase modulation across an incident beam's wave front. Using transverse electrodes (i.e. on the sides of thick wafers) require larger electrode separation and thus longer interaction lengths (for similar operating voltages) that can result in greater scattering (i.e. lower transmittance).

Both of these device configurations must also take into account ferroelectric hysteresis, photorefractive effect, depolarization due to scattering and saturation of the EO effect at high electric fields. The hysteresis of PLZT can be reduced by operation at temperatures of about 70-80° C. Photorefractive effect can be reduced by alternating the polarity of the driving voltage at approximately 50 Hz. However, both of these measures reduce the magnitude of the EO response of the material and result in a requirement for a higher operating voltage. High voltage operation results in greater scattering, depolarization and saturation effects. Device design to achieve the most efficient device configuration will need to find the optimal balance for all of these factors.

Design testing is greatly improved by using computer simulation to look at a multitude of parameter changes. We are using a finite element analysis program, in conjunction with rigorous material characterization that has allowed us to simulate arbitrary device configurations (see Appendix 10 and 11). We have been able to optimize for many individual parameters and are currently working on modeling that can optimize for multiple parameters to achieve optimal device design for MIN device implementation.

For narrow beam polarization rotation, we have found that a transverse electrode geometry provides the best combination of low voltage operation and uniform phase modulation. Using a push-pull configuration of MOSFET drivers (see Appendix 12), we have built a PLZT based polarization rotator capable of switching a square wave at 10 MHz. In Figure 7 we see that the PLZT rise time (top curve) of 30 ns is even faster than the driving signal (bottom curve). This faster rise time is due to a latency in the EO response of the material at low voltage. Experimental evidence shows that latency may be a function of frequency, device geometry and fabrication techniques.

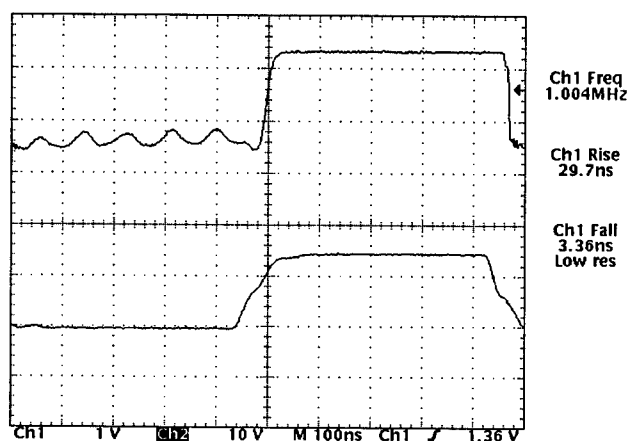


Figure 7. Bottom curve shows the 1 MHz driving signal of 160 V amplitude square wave with a rise time of approximately 100 ns and a fall time of 80 ns. Top curve shows the V_{π} response of the PLZT where the rise time is 30 ns and the fall time is less than 4 ns.

In order to understand the limits of the switching speeds we have also characterized PLZT response to an AC signal in terms of composition, applied voltage, temperature and device geometry. Experimental measurements of transverse electrode devices has shown that the EO effect exponentially decays to about 20% of the DC strength at approximately 10 MHz (see Figure 8). The two peaks at 3.1 and 6.5 MHz are amplified EO responses due to acousto-optic resonances. For applications that can take advantage of specific frequency range operation, device dimensions may be modified to provide significantly reduced operating voltages or enhanced EO response at proscribed frequencies due to this resonance amplification effect.

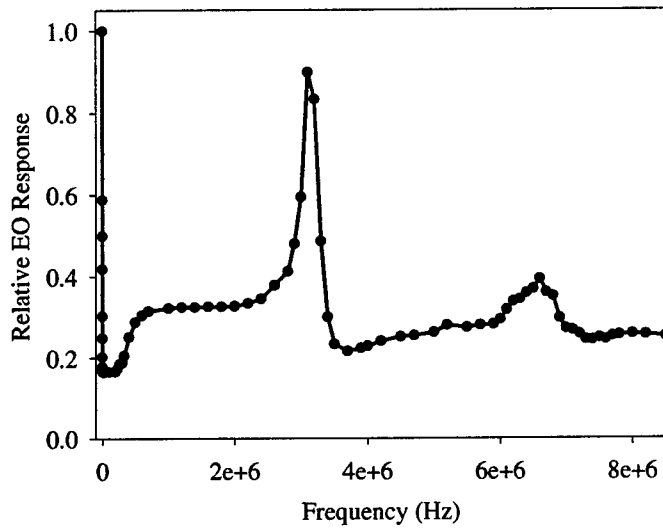


Figure 8. EO response of PLZT 8.8/65/35 from DC to 6 MHz with an applied ± 75 V signal amplitude. Peaks at 3.1 and 6.5 MHz are resonance points.

From DC to the 10 MHz range the EO response is primarily due to electrostrictive effects. At higher frequencies, ionic and electronic displacements are dominant. For applications that can withstand low contrast ratios or high driving voltages, it is expected that PLZT ceramic material will provide response times well below 1 ns. At UCSD we continue investigation of the high frequency EO response of PLZT.

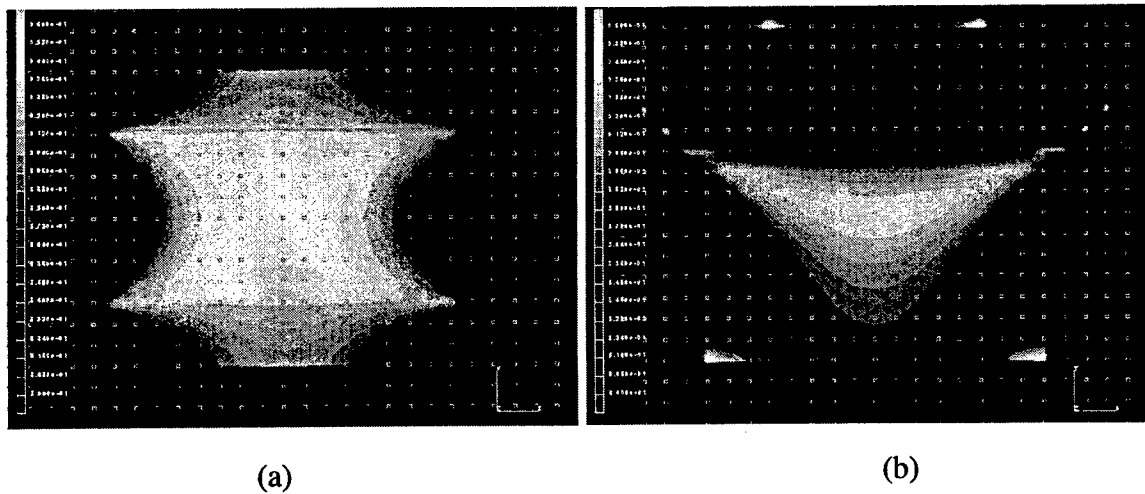


Figure 9. (a) Transmission intensity through crossed polarizers for single $150\ \mu\text{m}$ wide element, at V_π or the 'on' state, of array on $400\ \mu\text{m}$ pitch, where neighboring elements are also in the 'on' state. (b) Transmission intensity when neighboring elements are in 'off' state, i.e. 0 potential.

Implementation of an array of PLZT polarization rotators must take into account cross-talk between rotator elements. Cross-talk may be caused by electric field leakage between elements or electrostrictive or piezoelectric coupling of material properties. Using finite element analysis modeling we have investigated electric field distributions for arrays of rotator elements. Figure 9a and b show the simulated transmitted intensity for a single element in an array when neighboring elements are in the 'on' state and 'off' state, respectively. The intensity is directly related to the strength of the electric field within the PLZT material. It can be seen that the state of each element has a strong effect on the electric field distribution of neighboring elements. Experimental measurements on fabricated devices show similar behavior. Modification of device design allows for the significant reduction or elimination of measurable cross-talk.

6. System Demonstration

6.1. Folded Multistage MIN

To demonstrate a folded system, we designed and constructed an 8×8 optical MIN based on a fully connected Banyan architecture [1]. The design process of the system incorporates the following criteria: (i) maximization of the number of rings in off-axis Fresnel lenslets (ii) minimum feature size of diffractive elements conforming to available fabrication technologies (iii) separation of diffractive order beams. We developed a Gaussian beam analysis tool of the stable mode of the cavity that calculates these three parameters for a given cavity dimension.

The Gaussian beam spot size [19] at the BCGH plane was used as a limiting design constraint, since at that location the beam size is largest. For a spot size greater than the lenslet (i.e. array pitch), optical power would leak into adjacent elements giving rise to cross-talk. A spot size much smaller than the BCGH lenslet would result in diminished diffraction efficiency. Our pitch size was chosen to be 1 mm, dictated by dimensions of the pixelized polarization rotator device used (ferroelectric liquid crystal (FLC) device from DisplayTech, Model 10 \times 10B). A beam spot size of .825 mm was used, which provides for minimal cross-talk, high diffraction efficiency and power throughput (97% of the beam energy is contained in the 1 \times 1-mm square). Beginning with a cavity size of 200 mm, where the CGH array was placed in the center of the cavity (i.e. 100 mm from both mirror planes), the lenslet focal length is 50.4 mm with a 75 μ m

waist size at the mirror planes. The 1 mm pitch of the optical elements also dictates that the input light beam has an incidence angle of 0.5° . The polarization rotator array is placed adjacent to the CGH array to best match the 1 mm pitch of the FLC elements (see system schematic: Appendix 13).

The CGH array was fabricated in a quartz substrate using eight-phase levels, with diffraction efficiency greater than 90%. Each CGH element functions as an off-axis Fresnel lenslet, whose deflection angle is dictated by the interconnection pattern. For our 8×8 Banyan network the largest angle, corresponding to shifting the beam by three elements, is 0.015 radians. The BCGH element was fabricated using 4-phase level etching in YVO_4 crystal, selected for its high degree of birefringence, using the MOD approach. The BCGH elements were designed to use the 0 order (no phase encoding) for bypass mode and the +1 order (linear phase) for exchange mode. The diffraction efficiency was experimentally measured at 55.3% for the 0 order and 45.3% for the +1 order, with extinction ratios of 10:1 and 30:1, respectively. The patterned arrays of micro-mirrors were etched onto Al film evaporated on glass substrates with average measured reflectance of 92%. The circular micro-mirrors have diameters of $150 \mu\text{m}$, double the calculated beam diameter at the mirror planes (small enough to avoid the next diffraction order beam).

Experimental testing of this 8×8 MIN system was performed using a 488 nm CW Gaussian beam generated by an Ion Argon laser. Initial testing used one optical input channel modulated by a NEOS Model N71003 acousto-optic (AO) cell. The polarization state of the beam as it propagates through the network, which dictates the propagation path, is controlled by the DisplayTech 2-D array of FLC polarization rotators. Reconfiguration of the FLC elements is computer controlled with a minimum switching speed of 0.2 ms.

For 8×8 interconnectivity the beam reflects through the system 3 round trips (i.e. three layers of bypass-exchange switches). By diverting the beam after one or two passes we are able to use the same experimental system to test the performance of a 2×2 (single DBS switch) or 4×4 network. Figure 10 shows the output from a single DBS reconfiguring a single input, at a 1 kHz rate, between two output channels, where the input is a 50 kHz square wave signal (the relatively slow signal was used to allow simultaneous oscilloscope triggering of both frequencies). The top

trace is for the channel 1 signal and the bottom for the channel 2 signal. Both configurations give a $>20:1$ signal-to-noise ratio (SNR). Results using a 10 MHz signal show similar SNR, highlighting the optical transparency of the system. Testing of the complete 8×8 system resulted in a measured SNR of better than 10:1.

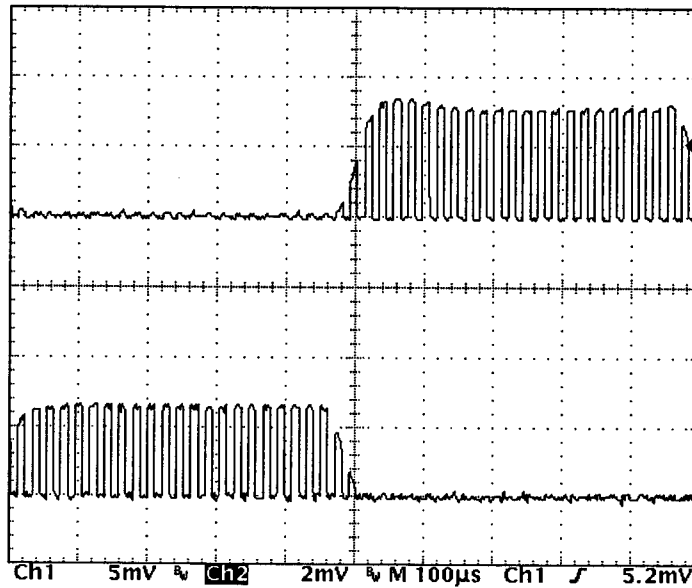


Figure 10. Folded DBS output shows 30:1 extinction ratio when switching signal between two channels. Rise and fall time of FLC reconfiguration is approximately 100 μ s.

Diffraction order	-4	-3	-2	-1	0	1	2	3	4
Percent power no phase encoding	3.5	1.2	9.1	1.4	55.3	1.5	2.8	0.9	5.0
Percent power linear phase encoding	1.0	8.0	1.3	22.3	5.8	45.3	1.3	2.3	0.8

Table 1. Linear phase BCGH element performance shows the diffraction efficiencies for several diffraction orders. The zero order contrast ration is less than 10:1.

The relatively low SNR for this system was primarily due to two factors: (i) Cross-talk from unwanted diffraction orders from the BCGH (Table 1 summarizes the efficiency per diffraction order) that were able to propagate within the system. The filtering of the micro-mirror pattern

was ineffective, since the mirror locations matched the higher diffraction orders in the design, allowing coupling of unwanted signals into other channels. (ii) Low contrast ratio, for the two polarization states, in the 0 order of the BCGH. In exchange mode, where each switch was designed to propagate only the +1 diffraction order, the residual 0 order component gives rise to strong cross-talk within the dilated switch.

6.2. Reduced Cross-talk MIN

Addressing the shortfalls of the first generation folded MIN, a redesign was initiated to decrease cross-talk. New BCGH elements were designed to use the +1 diffraction orders for bypass mode and the -1 orders for exchange mode. Also, the functionality of the lenslet array and the BCGH were combined to form a single, more complex optical element (see Figure 11).

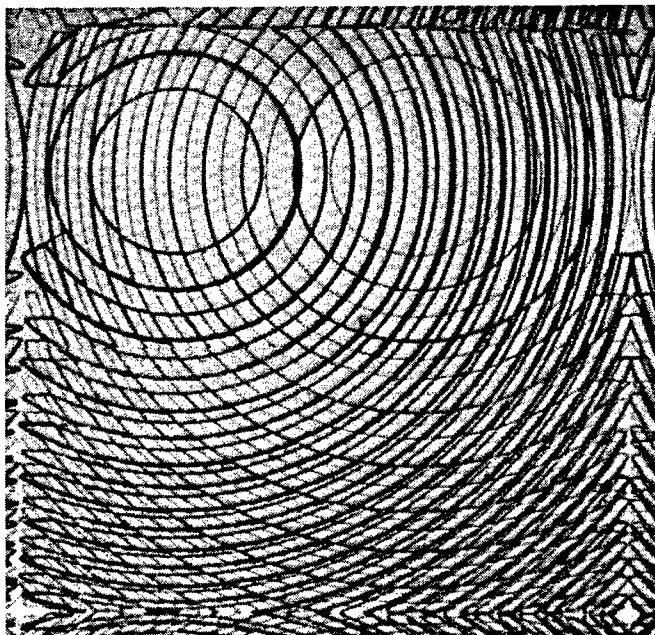


Figure 11. Microscopic photograph of new single BCGH element (part of an 8×6 array) where two Fresnel lenslets are encoded into a birefringent material so that each lenslet responds to one of two orthogonal polarization components.

This new design has the following advantages: (i) Only the high contrast ratio ± 1 orders are allowed to propagate within the resonator. The non-diffracted, 0 order light, which has relatively large residual component, is filtered out of the system. (ii) The unwanted higher order diffraction light is dispersed over a large area of the micro-mirror plane and it is *not* focused onto the micro-

mirrors (see Figure 12a). The amount of optical power incident on adjacent micro-mirrors (i.e. noise) and reflected back into the system is very small. It is determined by the ratio of the area of the micro-mirror to the area of the diffracted order at the plane of the mirrors (see Figure 12b). (iii) The number of components inside the resonator has been reduced to three: the polarization modulator, whose placement tolerance is relatively loose, the polarizer to filter out unwanted polarization components, and the BCGH element. This reduces losses due to reflection and further facilitates the compact packaging and simplified alignment of the system (see Figure 13).

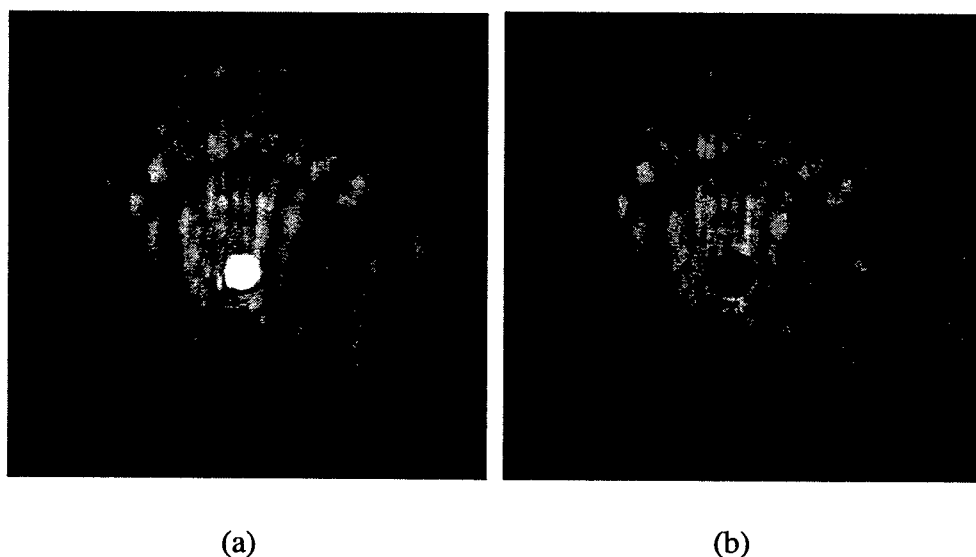


Figure 12. (a) Output of new combined BCGH element shows high diffraction orders dispersed over a large area. (b) Spatial filtering using micro-mirrors reflects only the wanted +1 or -1 order signals and allows noise to exit system.

Similar to the previous system design, the second generation design has three basic design criteria: (i) number of rings in the lenslets (ii) minimum DOE fabrication feature size and (iii) separation of diffractive order beams at the mirror planes. Choosing a minimum beam separation and DOE feature size, and varying the cavity dimensions to find the maximum number of lenslet rings (for high diffraction efficiency) optimizes the system design. With a minimum feature size of 10 μm and a maximum of 1 mm diameter lenslets, the new BCGH lenslets have $F/\#$ 50. Accounting for beam propagation through multiple optical elements (BCGH, rotator, and polarizer), the lenslet focal length was found to be 85.1 mm with a 300 μm waist size at one

mirror plane and $100\text{ }\mu\text{m}$ waist size at the other. The full resonant cavity length is 407 mm with the BCGH element placed 107 mm from the back side mirror. The circular micro-mirrors have diameters of $400\text{ }\mu\text{m}$ and $150\text{ }\mu\text{m}$, slightly larger than the calculated beam diameter at the mirror planes.

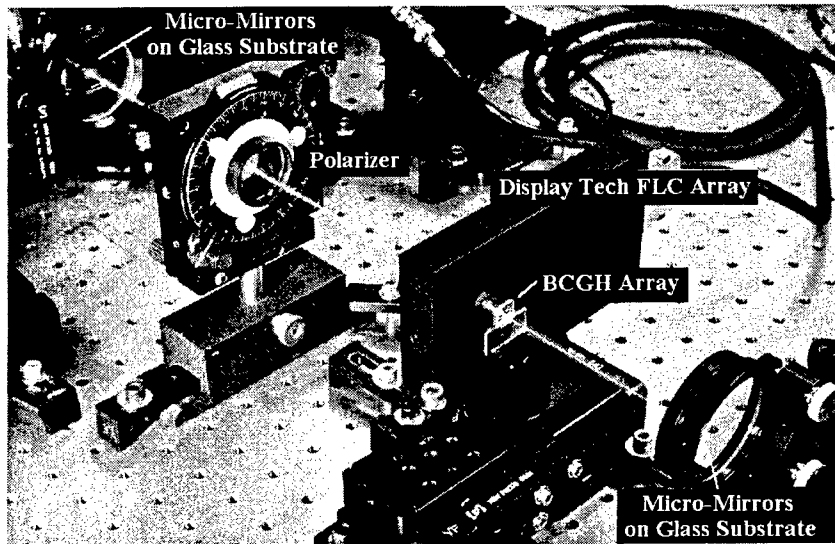


Figure 13. Picture of second generation folded MIN experimental system shows resonant cavity comprised of two glass substrates with Al micro-mirrors. Optical elements within the cavity are: (i) polarizer to filter linear cross-talk, (ii) FLC array for reconfiguration of interconnections and (iii) BCGH element (mounted on a glass slide) with polarization selective Fresnel lenslet encodings.

Figure 14 shows the output from a single 2×2 DBS reconfiguring between bypass and exchange mode at a 1 kHz rate. There are two input signals, one operating at DC and the other an AO modulated signal with a square wave at 20 kHz (this relatively slow input signal was used to allow simultaneous oscilloscope triggering of both AO and FLC reconfiguration frequencies). The top trace is for the channel 1 output and the bottom is for channel 2 output. Both configurations give extinction ratios (for one input signal, defined as the ratio between the 'on' state and 'off' state) greater than 59:1 and SNR (for multiple input signals, defined as the ratio of the signal to the noise at the same output, i.e. cross-talk) of greater than 57:1. Results using signals ranging from DC to 10 MHz show similar SNR, highlighting the optical transparency of the system. The extinction ratio for the DBS is significantly better than the individual BCGH elements, which shows how the DBS reduces cross-talk.

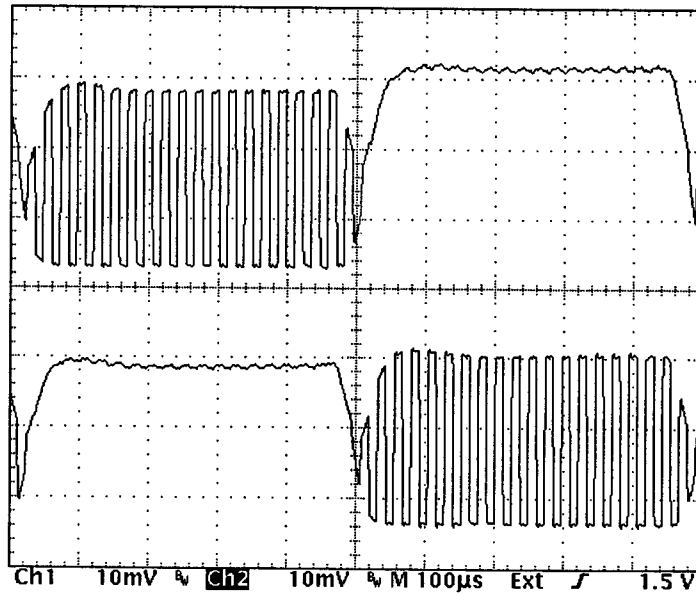


Figure 14. Output from DBS fabricated using new BCGH elements give 60:1 extinction and cross-talk ratios.

By allowing the beams to make two round trips through the cavity, a 4×4 MIN was experimentally characterized. Using a single DC input signal switching between four output channels, we measure an average SNR of 90:1 and extinction ratio of 120:1. To further investigate cross-talk we introduce a second input signal. The measured output amplitudes are shown in Figure 16. Most notable is that the output intensities vary dependent on the output as well as the input channel. This may be due to the variation of the Fresnel lenslet diffraction efficiency for different polarization states. The average minimal (i.e. the weakest output signal to the strongest output noise) SNR is 87:1.

The complete 8×8 interconnection system was characterized using a single DC input signal switching between all eight output channels. The output plane, where output signals may be coupled into optical fibers, was imaged onto a CCD camera (see Figure 16). The average measured SNR was better than 30:1. This relatively low SNR can be attributed to the strong background noise and small dynamic range of the CCD device used. We performed similar measurements using two input signals switching between all eight output channels that gave similar SNR results.

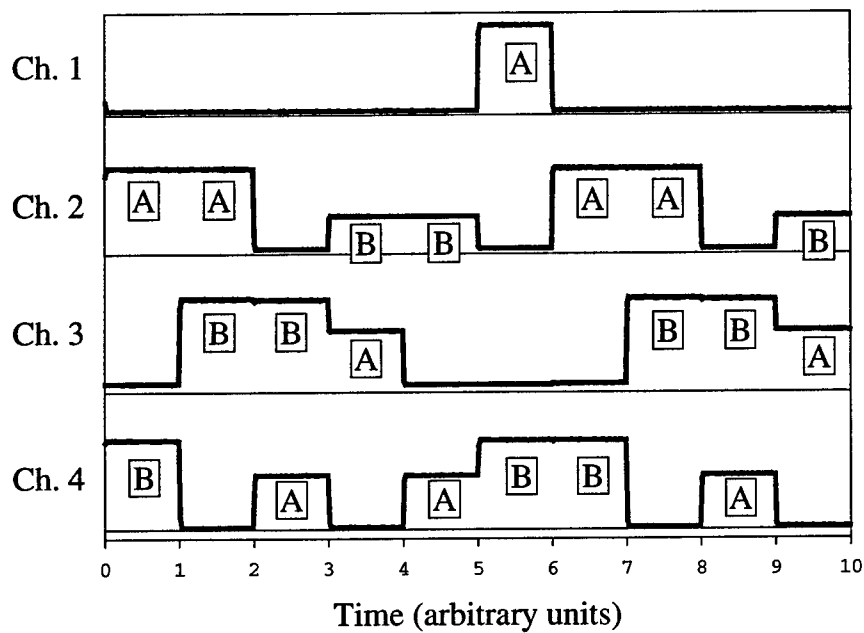


Figure 15. Output packets from 4x4 switch with two input signals. The two input signals (A and B) are routed to the four output channels by a host computer controller. Varying packet output intensity is due to a greater BCGH diffraction efficiency for the vertical polarization state (compared to the horizontal polarization state).

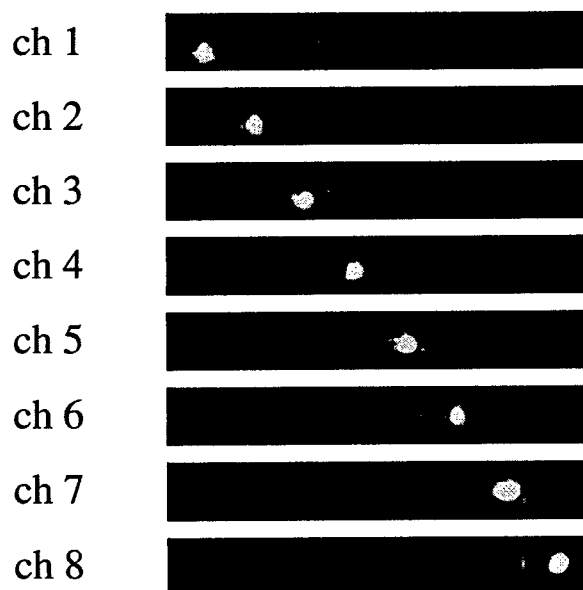


Figure 16. Switching a single DC input between eight output channels shows an average cross-talk ratio of 30:1.

7. Conclusions

The research results of this program can be summarized as follows:

(1) We have developed two new approaches to design and fabrication of birefringent computer generated holograms. The MOD approach, which can be fabricated on single substrates of birefringent materials, was designed, fabricated and experimentally tested. This design showed excellent diffraction efficiency and extinction ratios for the binary phase elements. However, due to ion beam etching inconsistencies, the performance of the four phase level elements was well below the efficiency predicted by scalar theory. The form birefringent FBCGH approach was designed using a combination of EMT and RCWA. A binary phase level device was fabricated and experimentally characterized. It has diffraction efficiencies higher than scalar theory predicts, which indicates that these elements have anti-reflection functionality.

(2) We have developed the Gated-Hold protocol for centralized control of blocking MIN. Unlike previous methods which have been analyzed by deterministic approaches, which doesn't allow a rigorous testing, the gated-hold protocol has been rigorously analyzed using stochastic methods. This protocol, in a cycle, will guarantee the delivery of all requests. The gated-hold protocol compares favorably to other blocking network protocols. However, the overhead imposed by the gated-hold protocol is characterized by the hardware complexity. Depending on the hardware configuration a different method of analysis may be required.

(3) We have designed and fabricated PLZT 9.0/65/35 based polarization rotation devices that can be switched at faster than 10 MHz with a 160V square wave driving voltage. This range of composition (8.8-9.65/65/35 PLZT) has been extensively characterized for frequency, temperature and geometry dependent EO response. The use of a novel transverse electrode geometry allows for coupling of acousto-optic effects for EO amplification at design frequencies. We have also developed a FEA based model of PLZT modulators, which successfully predicts device behavior and is used for device design and optimization.

(4) In this research we have designed and implemented an optical MIN using a novel folded optical DBS based on developed BCGH polarization selective technology. We have

demonstrated how this design allows for the elimination of the first-order cross-talk, ease of MIN system alignment and packability as well as filtering of unwanted high diffraction orders. The use of space variant lenslets also allows for the implementation of arbitrary MIN architectures. Comparison of ~30:1 extinction ratio for BCGH elements and ~60:1 extinction ratios for the DBS highlights the ability of the dilated switch to reduce cross-talk.

Our initial design of a folded DBS was based on switching between the 0 and +1 diffraction orders using simple linear phase encoded BCGH elements. However, this design allows the strong residual component of the 0-order light, with less than 10:1 contrast ratio, as well as high diffraction orders to propagate in system, which resulted in an average SNR of 10:1 for the 8×8 system. An improved design switches between the +1 and -1 diffraction orders and combines the space-variant lenslets and polarization selective elements. This new design results in an improved average SNR of 30:1 as well as further reducing the complexity of the system.

The four-phase level BCGH elements with low extinction ratios were the limiting factor in improving the SNR of the system. This poor performance can be attributed to inaccurate etching depths due to ion-etching device inconsistencies. Significantly higher SNR for the folded system can be expected with improved etching facilities and the use of alternative BCGH design approaches.

8. References

1. A.V. Krishnamoorthy, F. Xu, J.E. Ford and Y. Fainman, "Polarization-controlled multistage switch based on polarization-selective computer-generated holograms," *Appl. Opt.*, **36**(5) 1997.
2. P. Dietrich and R. Rao, "Dynamic Performance of a Circuit-Switched Interconnection Network," Presented at International Conference on High Performance Computing in New Delhi, India, December 1995.
3. P. Dietrich and R. Rao, "Request Resubmission in a Blocking, Circuit-Switched, Interconnection Network," *IEEE Trans. Comp.*, **45**(11) 1996..
4. F. Xu, R.C. Tyan, J. Ford and Y. Fainman, "Single Substrate Birefringent Computer Generated Holograms," *Opt. Lett.*, **21**(7) 1996.
5. F. Xu, R.C. Tyan, J. Ford and Y. Fainman, "Multiple Order Delay Holograms for Polarization and Color Selectivity," Presented at OSA Topical Meeting on Diffractive Optics, 1996.
6. F. Xu, R.C. Tyan, P.C. Sun, C.C. Cheng, A. Scherer and Y. Fainman, "Form-birefringent computer-generated holograms," *Opt. Lett.*, **21**(18) 1996.
7. R.C. Tyan, A.A. Salvekar, H.P. Chuo, C.C. Cheng, A. Scherer, P.C. Sun, F. Xu and Y. Fainman, "Design, fabrication, and characterization of form-birefringent multilayer polarizing beam splitter," *JOSA A*, **14**(7) 1997.
8. F. Xu, R.C. Tyan, P.C. Sun and Y. Fainman, "Fabrication, modeling, and characterization of form-birefringent nanostructures," *Opt. Lett.*, **20**(24) 1995.
9. R.C. Tyan, P.C. Sun and Y. Fainman, "Polarizing Beam Splitters Constructed of Form-Birefringent Multilayer Gratings," *SPIE Proceedings*, Vol. 2689, 1996.
10. P. Shames, P.C. Sun and Y. Fainman, "Modeling and Optimization of Electro-Optic Phase Modulator," *SPIE Proceedings*, Vol. 2693, 1996.

11. P. Shames, P.C. Sun and Y. Fainman, "Modeling Electric Field Induced Effects in PLZT EO Devices," *TOPS, Spatial Light Modulators*, Vol. 14, 1997.
12. D. Marom, P. Shames, F. Xu and Y. Fainman, 'Folded Free-Space Polarization-Controlled Multistage Interconnection Network,' Submitted to *Appl. Opt.*, January 1998.
13. F. Heismann, et.al., 'Polarization-independent photonic switching system using fast automatic polarization controllers,' *IEEE Phot. Tech. Lett.*, **5**(11) 1993.
14. G.A. DeBiase, 'Optical multistage interconnection networks for large-scale multiprocessor systems,' *Appl. Opt.*, **27**(10) 1988.
15. K.M. Johnson, M.R. Surette and J. Shamir, 'Optical interconnection network using polarization-based ferroelectric liquid crystal gates,' *Appl. Opt.*, **27**(9) 1988.
16. K. Noguchi, T. Sakano and T. Matsumoto, 'A rearrangeable multichannel free-space optical switch based on multistage network configuration,' *J. Light. Technol.*, **9**(12) 1991.
17. F. Xu, J.E. Ford and Y. Fainman, 'Polarization-selective computer-generated holograms: design, fabrication and applications,' *Appl. Opt.*, **34**(2) 1995.
18. K. Padmanabhan, and A. Netravali, 'Dilated networks for photonic switching,' *IEEE Trans. Commun.*, **35**(12) 1987.
19. A. Yariv, **Quantum Electronics**, Ch. 6, (John Wiley and Sons, New York) 1989.

Appendices

A. 1. Polarization-controlled multistage switch based on polarization-selective computer-generated holograms

A.V. Krishnamoorthy, F. Xu, J.E. Ford and Y. Fainman

Applied Optics, Vol. 36, No. 5, February 10, 1997

Polarization-controlled multistage switch based on polarization-selective computer-generated holograms

Ashok V. Krishnamoorthy, Fang Xu, Joseph E. Ford, and Yeshayahu Fainman

We describe a polarization-controlled free-space optical multistage interconnection network based on polarization-selective computer-generated holograms: optical elements that are capable of imposing arbitrary, independent phase functions on horizontally and vertically polarized monochromatic light. We investigate the design of a novel nonblocking space-division photonic switch architecture. The multistage-switch architecture uses a fan-out stage, a single stage of 2×2 switching elements, and a fan-in stage. The architecture is compatible with several control strategies that use 1×2 and 2×2 polarization-controlled switches to route the input light beams. One application of the switch is in a passive optical network in which data is optically transmitted through the switch with a time-of-flight delay but without optical-to-electrical conversions at each stage. We have built and characterized a proof-of-principle 4×4 free-space switching network using three cascaded stages of arrayed birefringent computer-generated holographic elements. Data modulated at 20 MHz/channel were transmitted through the network to demonstrate transparent operation. © 1997 Optical Society of America

1. Introduction and Background

There is a growing need in the telecommunications and data-communications industry for a scalable switch that can provide high-throughput communication between a large number of input-output (I/O) ports. Recent advances in the area of fiber amplifiers has spurred interest in transparent optical networks, wherein communication between users is achieved without multiple conversions between the optical and electrical domains.¹ The implementation of 16×16 and larger switches in a number of optical technologies is currently being pursued. Moreover, polarization compensators have been developed for single-mode fibers to permit automatic and stable control of the polarization state of output optical signals.² One might thus envision a switching system that uses polarization-dependent optical switches. Because of its low-delay, high-throughput

characteristics, such a switch may also find applications in a tightly coupled multiprocessor networking system or a parallel-processor-to-memory interconnection. In fact, polarization switching has been widely proposed for use in the context of free-space optical multistage interconnection networks³⁻⁷ for switch sizes up to 256×256 .⁸

In this paper we describe a novel free-space polarization-controlled optical switch design and present the implementation of a 4×4 photonic switch. The potential advantages of this design include no bulky birefringent optical components, fewer optical surfaces resulting in lower insertion loss, no path-dependent attenuation nonuniformity, a reversible optical path, and greater flexibility in choosing the optical interconnect topology and the resulting switch architecture. The switching system is based on a unique polarization-selective optical element capable of acting with an arbitrary independent phase function on illumination with horizontally or vertically polarized monochromatic light. This element, known as a birefringent computer-generated hologram (BCGH), is composed of two birefringent substrates etched with surface-relief patterns and joined face to face.

In Section 2 we review the BCGH technology and discuss its application to the basic 2×2 switch. In Section 3 we describe a new, nonblocking multistage-switch architecture that is well suited to a photonic

A. V. Krishnamoorthy and J. E. Ford are with the Advanced Photonics Research Department, Bell Laboratories, Lucent Technologies, Holmdel, New Jersey 07733. F. Xu and Y. Fainman are with the Electrical and Computer Engineering Department, University of California, San Diego, La Jolla, California 92093-0407.

Received 15 December 1995; revised manuscript received 18 September 1996.

0003-6935/97/050997-14\$10.00/0

© 1997 Optical Society of America

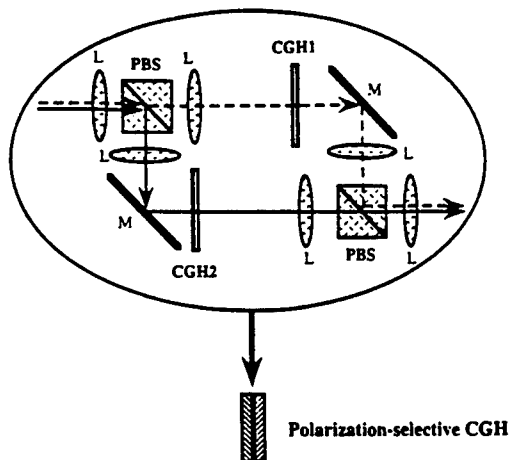


Fig. 1. Principle of operation of a polarization-selective hologram: The lenses represent two imaging stages of which the first imaging stage places one polarization on each Fourier-plane hologram and the second combines the outputs of the two polarizations. Here, 4- f imaging is required to transfer both the amplitude and phase of the incident wave front accurately. A BCGH combines the functionality of polarization beam splitters and associated interconnection optics in a single planar element.

implementation with BCGH technology. In Section 3 we also present a discussion of system trade-offs and a comparison of the architectures with other well-known multistage-switch designs. In Section 4 we present the implementation and characterization of a 4×4 photonic BCGH switch. In Section 5 we provide a summary and conclusions.

2. Application of Polarization-Selective Computer-Generated-Hologram Technology to a 2×2 Switch

A. Birefringent Computer-Generated-Hologram Technology

Multistage interconnection networks (MIN's) using polarization switching can be built with polarizing beam splitters (PBS's) and other bulk optics. However, system costs and complexity limit the number of stages and therefore the network size. It is possible to simplify the system substantially and eliminate many of the optical alignments by replacement of the discrete optical components with polarization-selective planar holograms (Fig. 1). A PBS, imaging lenses (L), and two computer-generated holograms (CGH's) can be replaced by a single polarization-selective CGH that has a different phase profile for each of the two orthogonal linear polarizations (Fig. 2). Polarization-selective holograms have been fabricated with various techniques, including optical recording of dichromated gelatin,^{9,10} organic media,¹¹ and photorefractive crystals,¹² as well as lithographic recording of polarization foil.¹³ However, we are primarily interested in a particular type of polarization-selective hologram—the birefringent CGH, that is, a CGH fabricated in birefringent media.¹⁴⁻¹⁷

A conventional kinoform CGH is a two-dimensional

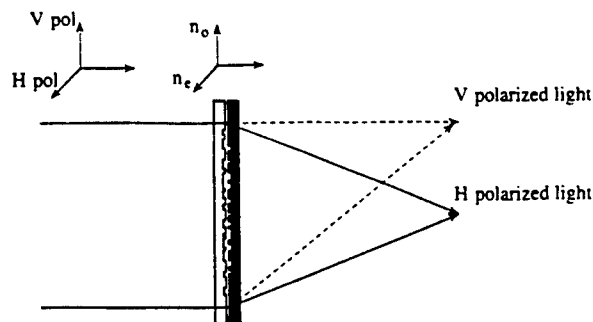


Fig. 2. Schematic diagram of the construction of BCGH's by the placement of two thin holographic elements face to face. At least one of the holograms is etched in an anisotropic medium (e.g., LiNO_3). H, horizontally; V, vertically; pol, polarized.

(2-D) phase profile that transforms the input light (e.g., a plane wave) into the desired output (e.g., an array of points). The desired continuous phase profile is first computed and then reduced to a minimum of data by pixellation, truncation to modulo 2π , and quantization into discrete values. This data array is then used to fabricate the hologram. A BCGH is similar in function to a conventional phase-only kinoform CGH, except that a BCGH has a different phase profile for each of the two orthogonal linear polarizations that illuminate the hologram. One fabricates kinoform phase-only CGH's by etching a surface-relief profile into an isotropic glass substrate. In a BCGH, the surface-relief profile is etched into a birefringent substrate. The birefringent substrate provides different indices of refraction, depending on the input polarization, that are used to differentiate between horizontally and vertically polarized inputs. The information content of the two arbitrary phase functions is contained in two etched substrates, joined face to face so that both profiles effectively lie in the same optical plane (see Fig. 2). These two substrates can apply an arbitrary phase for the two orthogonal linear polarizations.

The operation of the BCGH can be explained by consideration of the case in which one substrate is birefringent and the other is isotropic and in which the polarization of the incident light is aligned either along or perpendicular to the optical axis of the birefringent substrate. A ray transmitted through the birefringent substrate will have a different phase delay for each polarization because the indices of refraction differ. At each pixel, the phase angle between the two polarizations and the absolute phase delay of the rays depends on the thickness, hence the etch depth, of the birefringent substrate. This etch depth is chosen to produce the desired final phase angle between the two polarizations. The ray then passes through the isotropic substrate, where light of either polarization is delayed by the same phase angle, again depending on the etch depth. This etch depth is chosen to bring one polarization to the desired phase angle. Because the relative delay between polarizations is unaffected by the isotropic substrate,

the phase angles of both polarizations are simultaneously brought to the final desired values. The result is an optical element that can have high diffraction efficiency and provide arbitrary functionality for each of the two orthogonal linear polarizations of the input light.

The BCGH element effectively splits the input light beams by polarization, acts independently on each beam by use of separate CGH's, and recombines—redirects the output beams (see Fig. 1). The process of computing, etching, and assembling a BCGH is described in greater detail elsewhere.¹⁴⁻¹⁶ Methods of introducing artificial anisotropy in a transparent material by use of high-spatial-frequency, surface-relief nanostructures are also being investigated¹⁸; such investigation will ultimately permit BCGH elements to be fabricated on a single substrate.

B. Use of a Polarization-Selective Computer-Generated Hologram as an Optical 2×2 Switch

Two types of switch can be used for MIN's: 1×2 and 2×2 switches. A MIN can be made with $\mathcal{O}(\log_2 N)$ stages of N , 1×2 switches per stage and $2N$ links between stages.¹⁹ Switching is achieved by the choice of the output link each input takes. A switching MIN can also be constructed with $\mathcal{O}(\log_2 N)$ stages of $N/2$, 2×2 switches per stage and N links between stages. Switching is achieved by the choice of the state of each 2×2 switch. This latter type of network is the one pursued in this paper.

As shown in Figs. 3(a) and 3(b), it is often convenient to build a 2×2 switch by use of 1×2 switches.¹⁹⁻²¹ Figure 3(c) illustrates the allowed and disallowed states of the switch when 1×2 switches with passive combining are used to generate a 2×2 switch. The disallowed configurations of a 2×2 switch correspond to both inputs accessing the same output. For the case of a BCGH 2×2 switch, this situation would correspond to both polarizations having the same deflection angle at the output. If the inputs to the 2×2 switch have orthogonal polarizations, then the outputs will also have orthogonal polarizations when the axes of the electro-optic polarization modulator are aligned so as either to pass both polarizations without change or to rotate the polarizations of *both* beams by 90° . Hence, the disallowed configurations can be avoided if one ensures that the inputs to a 2×2 switch have orthogonal polarizations and if a 0° or 90° polarization-rotating switch is used.

A 2×2 polarization switch will thus require two BCGH planes and a polarization-rotator plane, as shown in Fig. 4. The two inputs are both directed into the first BCGH, which combines and focuses the two inputs into a polarization rotator (PR). After being combined, the two modulated beams propagate in the same direction through the PR; this step is essential to obtaining high-contrast modulation because polarization rotators are strongly angle sensitive. The PR sandwiched between the two BCGH elements controls the state of the 2×2 switch, i.e., either a crossover or straight-through connection. If

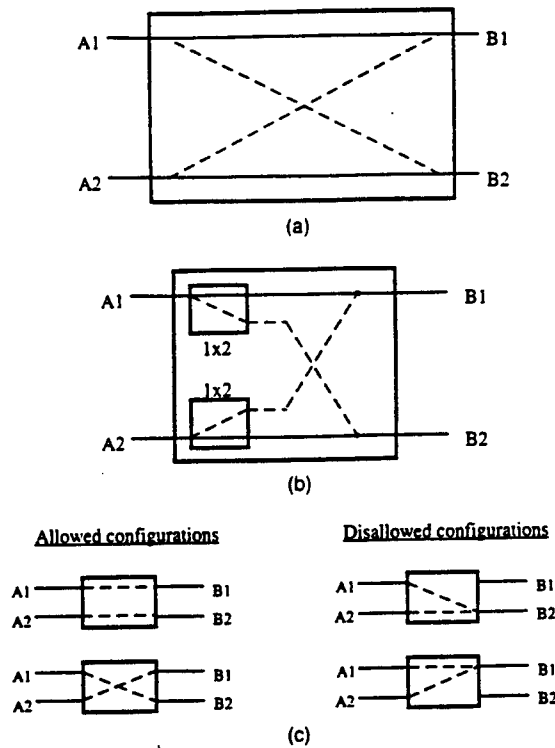


Fig. 3. Fabrication method and states of 2×2 switches: (a) A 2×2 switch either passes the inputs straight through or exchanges the inputs. (b) A 2×2 switch can be implemented by use of two 1×2 inputs with their respective outputs tied together. (c) Allowed and disallowed states. Disallowed states must be carefully avoided. In a BCGH implementation this is ensured by the requirement that the two inputs have orthogonal polarizations and by use of a 0° or 90° PR switch.

the PR is in the OFF state, the two beams propagate straight through, maintaining their original polarizations. When the PR is turned ON, the two beams exchange polarizations.

After transmission through (and possibly modulation by) the PR, the second BCGH element deflects the beams into two different directions, depending on their polarization states. Figure 4 shows the beam

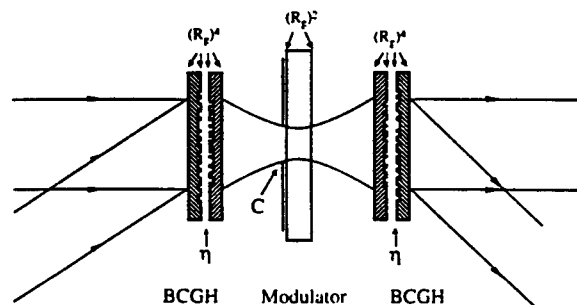


Fig. 4. Components of a 2×2 switch: two BCGH's and a polarization modulator. η is the switch efficiency, R is the transmittance, and C is the coupling efficiency associated with clipping losses, which are incurred when imaging a beam through a modulator aperture.

focused through a small aperture, which may or may not be necessary, depending on the modulator technology. Note that a BCGH switching element combines the functionality of a 2×2 switch and the associated holographic interconnection optics before and after the 2×2 switch. A 2×2 BCGH switch also has the same functionality for beams propagating through the optics in reverse, in principle making the network path reversible.

C. Insertion loss calculation for a 2×2 Polarization Switch

Optical loss in a BCGH switch is due to reflections from the dielectric surfaces and the diffraction efficiency of the holograms. To increase the diffraction efficiency of a BCGH hologram, and thus reduce the insertion loss (attenuation) of a BCGH-based switch, multilevel phase BCGH's are necessary. The theoretical diffraction efficiency of a CGH is proportional to the number of phase levels Φ used in its construction.²² The efficiency of a BCGH η_{BCGH} is the product of the efficiencies of its two component holograms:

$$\eta_{\text{BCGH}} = \text{sinc}^2\left(\frac{1}{\Phi_a}\right) \text{sinc}^2\left(\frac{1}{\Phi_b}\right). \quad (1)$$

For instance, the use of four phase levels ($\Phi_a = \Phi_b = 4$) in each of the holographic elements results in an optimum BCGH diffraction efficiency of approximately 64%. Several four-level phase BCGH elements, each consisting of a 4×4 array of blazed gratings, were designed and fabricated for application to the MIN described below in Section 4. The grating periods were $40 \mu\text{m}$, and the smallest feature size in the hologram was $10 \mu\text{m}$. A diffraction efficiency of 26% and a polarization contrast ratio of 130:1 were measured for the four-level phase element. These values can be compared with a measured diffraction efficiency of 12% and a contrast ratio of approximately 50:1 with a binary phase element having the same feature size. By tilting the hologram to compensate for alignment errors, Xu *et al.*¹⁶ have achieved best-case results for a four-level phase hologram of a 60% diffraction efficiency and a 160:1 contrast.

Surface-reflection losses at the BCGH substrate R_g and the modulator substrate R_m contribute to the switch loss. A clipping loss C occurs when the beams are focused through an aperture at the modulator. The total switch efficiency is then

$$\eta_{\text{switch}} = \eta_{\text{BCGH}}^2 R_g^8 R_m^2 C. \quad (2)$$

If we assume that each optical surface is antireflection coated with a single dielectric layer (to permit the maximum range of input angles) and that 16-level phase holograms are used, then these constants can be estimated to be $\eta = 98.4\%$, $R_g = R_m = 99\%$, and $C = 98.6\%$. The total switch efficiency is then 86.3%, producing an insertion loss of $10 \log_{10}(\eta_{\text{switch}}) = -0.638 \text{ dB}$. The calculations presented here represent the switch performance of a

transmission device. Note that BCGH components may also be used in conjunction with smart-pixel technology. Depending on the particular device technology, this combination can introduce other surfaces (e.g., a common substrate that holds the circuit and the modulator materials).

If the polarization rotation were exactly $\pi/2$ and the BCGH's distinguished completely between the two polarizations, the cross talk would be zero and the signal-to-noise ratio (SNR) of a 2×2 switch would be infinite. In practice, one can define the cross talk coming from one switch to be δ_c and the maximum number of switches in one path to be S . Then the SNR of the entire network is given by

$$\text{SNR}_{\text{network}} = \log_{10}(1/\delta_c) - \log_{10} S. \quad (3)$$

Note that δ_c is a critical factor that will determine the choice of architecture and maximum network size. In this paper we consider two examples: $\delta_c = 0.1\%$ and $\delta_c = 1\%$. These two cases are typical of currently achievable technology for bulk and pixellated BCGH switching elements.

3. Architecture of the Birefringent Computer-Generated Hologram Multistage Switch

A. Switch Architecture

The Stretch network is a class of self-routing MIN's that provides a continuous performance-cost trade-off between two of its extreme forms: the fully connected space-division switch (or crossbar switch) and the Banyan multistage network. Stretch networks utilize a destination-tag-based routing algorithm; that is, for each input channel, the necessary I/O path through the network can be determined on a stage-by-stage basis solely from the specified destination of the input packet. Stretch networks can be designed to achieve a low delay and arbitrarily low blocking probabilities for various traffic patterns without using internal buffers in the switching fabric. A common feature of all Stretch networks is that each stage in the multistage switching network uses a simple perfect-shuffle interconnection or any of the topologically equivalent connection patterns.²³ In this paper we are concerned with a specific nonblocking Stretch network with N I/O channels and k -shuffle interconnection between stages. The broader class of Stretch networks is described in more detail elsewhere.²⁴

An example of the switch architecture for $N = 8$ channels is presented in Fig. 5(a). In this network, the fan-out (splitting) stages [Fig. 5(b)] permit partial contention-free routing of the first $\log_2(N - 1)$ bits of the destination address for each of the N inputs; the switching stage provides the routing on the last bit of the destination address, and the fan-in (combining) stages [Fig. 5(c)] concentrate the outgoing data. The fan-out and fan-in stages provide contention-free demultiplexing and multiplexing, respectively, of each input signal. The fan-out stage is connected to the switching stage by use of a two-shuffle connection

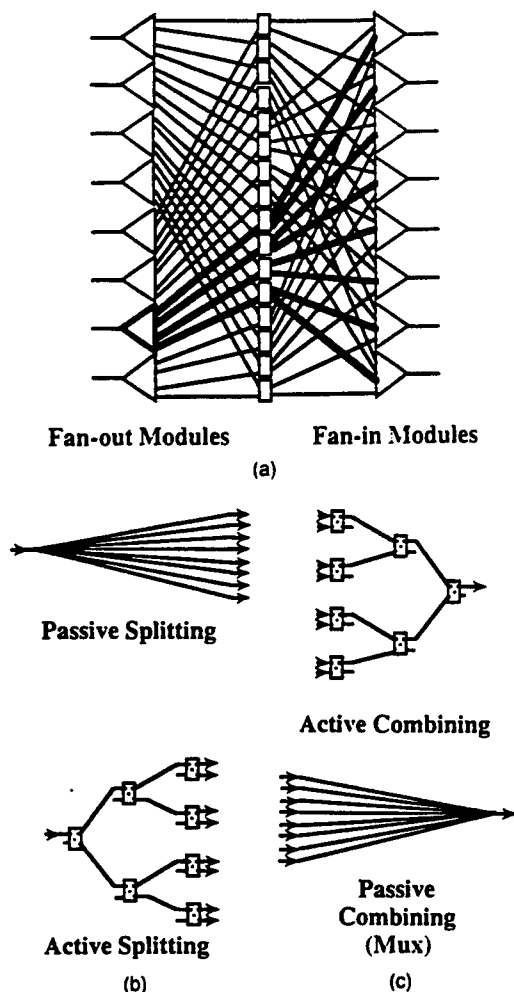


Fig. 5. (a) Stretch switch with eight inputs-outputs, a fan-out of 4, and one stage of 2×2 switches. The switch is nonblocking, and the first-order cross talk of the network is equal to the cross talk of a single switch. All lines represent point-to-point connections. (b) Fan-out modules may either be passive or active. (c) Fan-in, similarly to fan-out, may be active, passive with optical fan-in, or implemented with separate detectors and electrical multiplexing (Mux).

pattern, and the switching stage is connected to the fan-in stage by use of a $(N/2)$ -shuffle connection pattern. Notice that the fan-out for each input is $N/2$, which is half the fan-out for a fully connected switch. A single stage of 2×2 switches is used in the center of the switching fabric. An important consequence is that the switch is strictly nonblocking, with a unique constant-length path from each input port to each output port. This can be verified visually from Fig. 5(a) and is proved in Ref. 25. The highlighted connections in Fig. 5(a) show the path from a specific input to each of the output ports.

This nonblocking architecture is well suited for parallel optical implementation because it uses a single stage of $N^2/2$, 2×2 switching elements and because the destination-tag-routing property of the multistage switch results in a simple path-hunt al-

gorithm that can readily be accomplished in parallel if required. The structure of this switch is, in principle, similar to the extended generalized shuffle network described in Ref. 20, except that the Stretch network has an exact multiple of $\log N$ logical stages (including the fan-out modules) between the input and output ports, thereby ensuring a self-routing structure, hence a simple routing algorithm, that may be applied to each channel independently of the others.

The implementation of the fan-out and fan-in stages in an optical Stretch network is critical to the network's performance. The fan-out stage may be passive (i.e., optical broadcast) [passive splitting (PS)], which results in a maximum $2/N$ power efficiency, or it can be active [active splitting (AS)] i.e., built by use of a tree-based architecture with $N/2$ additional 1×2 switches per fan-out module.²⁶ Similarly the fan-in stage can be active (built with 2×1 switches) [active combining (AC)], passive with optical fan-in [passive combining (PC)], or can use $N/2$ separate receivers per output module with electronic multiplexing. For the active tree-based fan-out and fan-in modules, the control lines in a stage are typically tied together for convenience; hence each module will require $\log_2(N - 1)$ control lines to control $N/2$ switches.

When AS is used, each fan-out stage consumes the first $\log_2(N - 1)$ bits of the destination address of the corresponding inputs, and the 2×2 switching stage consumes the last bits to achieve a unique output address for each of the N inputs. In this mode, the network can be self-routing. The nonblocking network structure ensures that no permutation can result in network blocking or a disallowed switch state. If PS and AC are used, then the network controller must work in reverse by use of sender-tag routing, in which the fan-in unit is set according to the last $\log_2(N - 1)$ bits of the sender address and the 2×2 switching stage is set according to the first bit of the sender tag. In addition, the top half of the inputs must be polarized orthogonally to the lower half of the inputs to ensure proper operation of the 2×2 BCGH switches. When active fan-out and active fan-in modules are present, both destination-tag and sender-tag algorithms are used. In all cases, the path-hunt algorithm and the switch-setting process may be performed independently and in parallel for each of the N channels. This property enables the path hunt to be performed in $O(\log N)$ time steps. Note that special care must be taken to ensure that the inputs to a 2×2 switch in the switching stage have orthogonal polarizations. When PS is used together with a shuffle interconnection topology, then one method of ensuring this is to polarize the top half of the inputs orthogonal to the lower half. If AS is used with an optical shuffle, then the polarizations of subsequent outputs alternate, and the polarizations of the lower half of the inputs mirror the upper half.

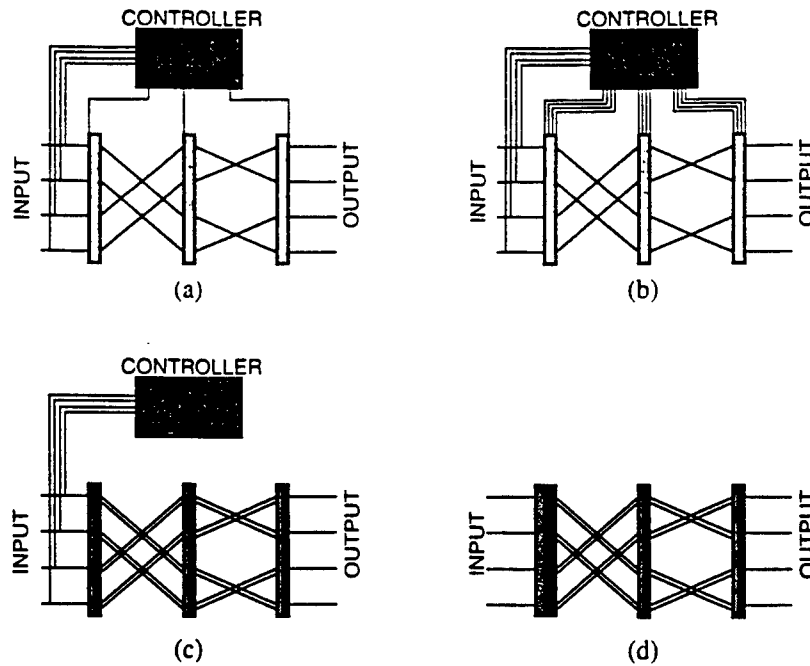


Fig. 6. Possible control algorithms for BCGH-based MINs: (a) centralized control with global switching, (b) centralized control with direct injection, (c) centralized control with packet headers, and (d) distributed control with self-routing headers.

B. Control Algorithms

In a BCGH-based Stretch network, the switch can operate in a transparent circuit-switched mode and can be either locally or globally routed. The setup and reconfiguration times depend on the specific polarization-modulator technology, but after the network state is set the data-transmission rate is limited only by signal attenuation, cross talk, and factors external to the multistage switch, i.e., the transmitter and receiver responses. Communication involves three phases: (i) circuit establishment, in which end-to-end circuits are set up before transmission begins; (ii) data transmission, in which the data-modulation rate is decoupled from the reconfiguration of the network; and (iii) circuit disconnect. The switch described in this paper is of the space-division switching type, with the added possibility that the inputs to the network may be time-division or wavelength-division multiplexed (e.g., from an optical-fiber bundle).

A key concern is the control of the network's switching states. Several distinct types of control algorithms can be defined: centralized control with global switching, centralized control with direct injection, centralized control with packet headers, and distributed control with self-routing packet headers.²⁷ Figures 6(a)–6(d) show each approach schematically. In centralized control with global switching, the switches in each layer of the network are linked and can only switch as a unit. The number of control lines is greatly reduced, but only a single arbitrary interconnection of one input to one output can be made at a time. This functionality is

useful in the fan-out (fan-in) modules for which a single input (output) must be steered to a select channel, so that the control lines of the 1×2 (2×1) switches in a column may be tied together. This reduction in control lines potentially increases second-order cross talk through the network.

The second method is to use centralized control with direct injection. In this scheme the routing algorithm is calculated at a central controller that determines switch settings and accesses the switching elements sequentially or in a semiparallel fashion. For controlling large networks it becomes essential to have an architecture, such as the Stretch network, that allows path hunts to be performed in parallel.

Another approach is to use centralized control with packet headers. The routing algorithm is performed at a centralized controller, but the process of setting the individual switches of the MIN is implemented by use of packet headers that propagate through the network. This approach can be achieved by use of ≤ 30 transistors per switch and can be achieved by use of a BCGH coupled with smart-pixel technology. The routing information must still be distributed to the first-stage array, but not to the arrays at each layer of the network. For BCGH-based networks using this form of control, virtual circuit switching is achieved by the specification of a dedicated time interval for the packet headers with control information to propagate through the network and set up the required data paths. As soon as the switches have been set, passive transmission (no detection and rebroadcast) at high data rates is possible. For transparent operation, control may optionally be at a

distinct wavelength or on a separate path. The main advantage of this method is a reduced controller pin-out (a factor of at least $\log_2 N$ fewer than direct injection). However, this approach requires synchronous network operation and a technology that can provide some intelligence at each pixel location.

A fourth approach is to use distributed control with self-routing packet headers. Here, no central controller is needed. This algorithm is typically associated with fast packet switching. A smart-pixel implementation will require 50–75 transistors per 1×2 switch.²¹ This approach uses the same smart-pixel hardware and interconnection topology as the centralized approach with packet headers. These packet-header control approaches are relevant for only BCGH networks that use active fan-out or active fan-in (or both).

C. Comparison of Nonblocking Networks

Two key performance metrics for a passive multistage photonic switch architecture are its optical attenuation and its SNR or cross talk. These in turn are related to the amount of optical fan-out per input, the number of stages in the multistage-switching network, and the number of switches an input signal must traverse to reach its intended output (path length). In the following we analyze the performance of the nonblocking Stretch network in terms of its cross talk and attenuation relative to some well-known nonblocking MIN architectures that are suitable for implementation with the 2×2 BCGH switches. These include the crossbar, N -stage planar,²⁸ Benes,²⁹ dilated Benes,³⁰ and Batcher-Banyan networks.³¹ A more detailed discussion of these switching-network architectures can be found in Refs. 28–32.

In terms of nonblocking architectures for photonic switching, the most well-known switch architecture is the crossbar. The crossbar (or full space-division switch) is a strictly nonblocking architecture with N^2 switches and a worst-case path length of $2N - 1$. In a crossbar, the path length and the signal skew depend on the specific interconnection path. The N -stage planar network is a rearrangeable nonblocking architecture requiring N stages with $N/2$ switching elements per stage. It evolved from the crossbar, providing fewer switches and a planar (no crossover) architecture by use of only the nearest-neighbor interconnection and equal path lengths. The total number of switches is $N(N - 1)/2$, and the maximum path length is N .

In terms of nonblocking multistage architectures that require significantly fewer 2×2 switches, a widely studied architecture is the Benes network. The rearrangeable nonblocking Benes network consists of two $\log_2 N$ networks placed end to end. The network has $2 \log_2 N - 1$ stages, which is the theoretical minimum number of stages required for rearrangeable nonblocking operation. This network provides an equal path length, low latency, and low switch count at the expense of a more complicated routing overhead.

The dilated Benes architecture was a modification of the Benes network designed to remove effects of cross talk that plague architectures with a $\log N$ or greater depth. This is a Benes network that has been doubled in width while the initial number of inputs and outputs has been maintained. Dilated Benes networks thus have $2 \log_2 N$ stages and N switches per stage. The network has the unique advantage that no switching element carries more than one active signal. Hence, first-order cross talk is eliminated. Optical cross talk from another channel can be mixed with an active signal only by its passing through two nominally off switches. When this second-order cross talk is low, the network can achieve a large SNR. Finally, the Batcher-Banyan network is a self-routing network consisting of a sorting network followed by a routing network. The total number of stages is equal to $(1/2)\log_2^2 N + (3/2)\log_2 N$.

Table 1 contains a summary of the attenuation, SNR, number of stages, and total number of switches for each of the nonblocking architectures described above versus the network size N . Table 2 similarly shows the scaling of the SNR, attenuation, and switch count of the various configurations of the Stretch network. The results depend strongly on the type of fan-out and fan-in modules used. For instance, the number of stages in a Stretch network depends on the design of the fan-out (splitting) and fan-in (combining) stages. It is $\log N$ if either the splitting or the combining is active (AS/PC or PS/AC), $2 \log_2 N - 1$ if both the splitting and combining stages are active (AS/AC); or 1 if no active switching is used (PS/PC with separate receivers).

As a result of SNR degradation, optical fan-in to a common detector is feasible only for smaller networks. The first-order cross-talk isolation of the AS/AC Stretch network is equal to the cross-talk isolation of a single switch, independent of network size. The second-order cross talk is much smaller in magnitude than the first-order cross talk and can be neglected. The attenuation and the SNR for $\beta_c = 30, 20$ dB, respectively, are graphed in Figs. 7–9. The corresponding performance of several Stretch networks is shown for comparison. The dotted cutoff lines show the maximum achievable sizes of each architecture, assuming a maximum acceptable attenuation to be 30 dB (99.9%) and the minimum SNR to be 11 dB (corresponding to a bit error rate of 10^{-9}).³² It should be noted that the SNR can be increased at the price of increased attenuation. For example, if the fan-out in the Stretch network were increased from $N/2$ to N , there would be no first-order cross talk and the SNR could be doubled. In this case second-order cross talk must be accounted for. The resulting network would be identical to a full space-division switch,²⁸ and it would have increased attenuation and would also require more switching elements (Table 2).

It is evident that nonblocking networks, such as crossbars or N -stage planar networks, are not well suited to large-scale implementation with a BCGH.

Table 1. Performance Scaling for Several Well-Known Photonic Switch Architectures in Terms of the Network Size N

Architecture ^a (N inputs/ N outputs)	Attenuation ^b (dB)	SNR ^c (dB)	Number of Logical Switching Stages	Number of 2×2 Switches ^d
Crossbar (nonblocking)	$(2N - 1)\alpha_s$	$\beta_s - 10 \log_{10}(N - 1)$	N	N^2
N -stage planar (nonblocking)	$N\alpha_s$	$\beta_s - 10 \log_{10} N$	N	$N(N - 1)/2$
Batcher-Banyon (nonblocking)	$[1/2 \log_2 N (\log_2 N - 1) + \log_2 N]\alpha_s$	$\beta_s - 10 \log_{10}(\text{number of stages})$	$1/2 \log_2 N (\log_2 N - 1) + \log_2 N$	$N/2(\text{number of switches})$
Benes (rearrangeable, nonblocking)	$(2 \log_2 N - 1)\alpha_s$	$\beta_s - 10 \log_{10}(2 \log_2 N - 1)$	$2 \log_2 N - 1$	$N/2(2 \log_2 N - 1)$
Dilated Benes (rearrangeable, nonblocking) ^e	$2(\log_2 N)\alpha_s$	$2\beta_s - 10 \log_{10}(2 \log_2 N - 1)$	$2 \log_2 N$	$2N \log_2 N$

^aThe architecture was circuit switched.^bFor the worst-case optical path loss. α_s is the insertion loss per switch, in decibels.^cFor the worst-case SNR. β_s is the cross-talk isolation per switch, in decibels.^dIn the entire network.^eThe SNR in this case is due to second-order cross talk.

In terms of attenuation limits, the Batcher-Banyan MIN scales up to 256 I/O ports. The Benes, dilated Benes, and Stretch networks with AS scale well beyond $N = 1000$, making them good choices (Fig. 7). Power losses that are due to fan-out limit Stretch networks with PS to fewer than 512 I/O ports. When the cross-talk isolation of a switch β_s equals 30 db, all these networks perform well in terms of SNR. However, when β_s is lowered to 20 db, the cross talk from the switches along the routing paths severely limits the scalability of Benes networks, and to a lesser extent the Stretch networks with PS. In this

case, either a dilated Benes network or a Stretch network with AS must be used to counter the deleterious effects of cross talk (Figs. 8 and 9).

The conclusion is that the nonblocking Stretch network with AC is a suitable candidate for a BCGH-based switch implementation and has good potential for extension to large networks. The main advantages of the Stretch network over other suitable multistage networks, such as the dilated Benes, are its nonblocking operation without the need for rearrangement and its simple, parallel path-hunt capability. Among the nonblocking networks, it has the

Table 2. Performance Scaling for Various Configurations of the Stretch Network versus the Network Size N^a

Architecture ^b (N Inputs/ N Outputs)	Attenuation ^c (dB)	SNR ^d (dB)	Number of Logical Switching Stages	Switches	
				Number ^e	Type
Stretch AS/AC tied control lines	$(2 \log_2 N - 1)\alpha_s$	β_s	$2 \log_2 N - 1$	$N(N - 2) + (N^2/4)$	1×2 2×2
Stretch AS/AC separate control lines	$(2 \log_2 N - 1)\alpha_s$	β_s	$2 \log_2 N - 1$	$N(N - 2) + (N^2/4)$	1×2 2×2
Stretch PS/AC tied control lines	$(\log_2 N)\alpha_s + 3(\log_2 N - 1)$	$\beta_s - 10 \log_{10}(\log_2 N - 1)$	$\log_2 N$	$N(N/2 - 1) + (N^2/4)$	1×2 2×2
Stretch AS/PC tied control lines; fan-in to receivers	$(\log_2 N)\alpha_s$	$\beta_s - 10 \log_{10}(N/2)$	$\log_2 N$	$N(N/2 - 1) + (N^2/4)$	1×2 2×2
Stretch PS/PC separate receivers	$\alpha_s + 3(\log_2 N - 1)$	β_s	1	$(N^2/4)$	2×2
Full space-division switch AS/AC fan-out equals N	$(2 \log_2 N)\alpha_s$	$2\beta_s - 10 \log_{10}(\log_2 N)$ (second-order cross talk)	$2 \log_2 N$	$2N(N - 1)$	1×2

^aEither AS or PS may be used in the fan-out module. For small networks, the fan-in modules may use PC with either optical fan-in or separate receivers. For large networks, the fan-in modules should use AC. The control signals in a stage of an active splitting (fan-out) or an active combining (fan-in) module can be tied together to reduce the number of separate control lines.

^bThe architecture is circuit switched.^cFor the worst-case optical path loss. α_s is the insertion loss per switch, in decibels.^dFor the worst-case SNR. β_s is the cross-talk isolation per switch, in decibels. SNR is limited by first-order cross talk unless otherwise noted.^eNumber in the entire network.

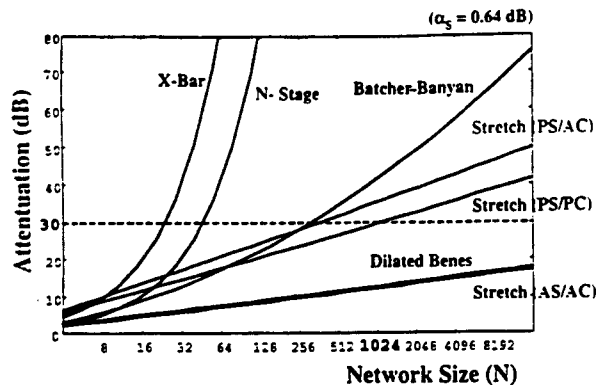


Fig. 7. Network attenuation α_s (in decibels) versus the number of input ports N for BCGH-based MIN's with the assumption of a 2×2 switch insertion loss of 0.64 dB. X-Bar, crossbar.

fewest number of optical stages. This reduces the network delay and control complexity. The penalties are its increased attenuation when PS is used and its larger number of switches when AS is used. For switching networks with less than 512 I/O ports, PS can be used; for larger networks, AS may become necessary. Technological limits to the BCGH Stretch network with AS arise primarily from array-size limits of the pixellated polarization modulator ($N^2/4$ pixels needed) and the maximum size of the birefringent hologram.

4. 4×4 Switch Demonstration

BCGH components have previously been evaluated for a 1×2 switch¹⁵ and a 2×2 switch.^{16,33} Here we describe the first multistage switching network demonstration based on cascaded arrays of polarization-selective holographic components. We have designed and implemented a three-stage, 4×4 BCGH optical multistage switch that can be scaled to larger sizes. The 4×4 BCGH-Stretch network used centralized control with global switching for the fan-out module and centralized control with direct injection for the 2×2 switching stage. Figure 10

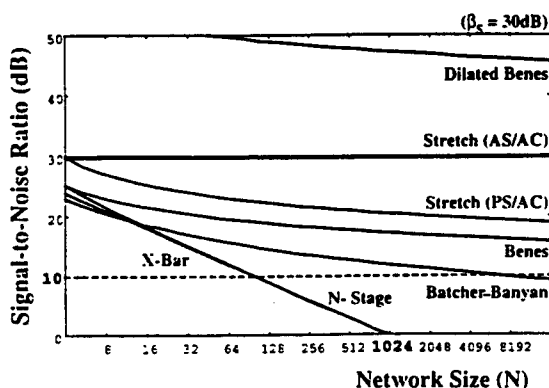


Fig. 8. SNR β_s (in decibels) versus the number of input ports N for BCGH-based MIN's with the assumption of a 2×2 switch SNR of 30 db. X-Bar, crossbar.

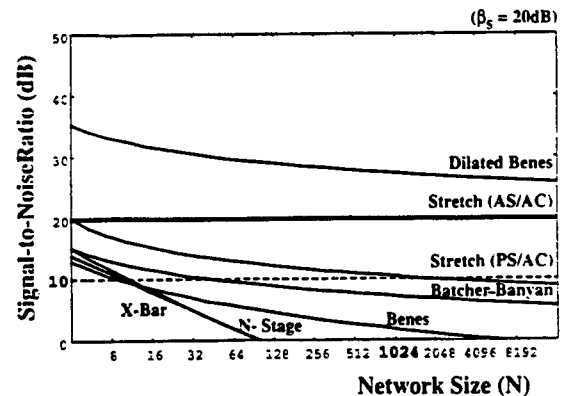


Fig. 9. Network SNR β_s (in decibels) versus the number of input ports N for BCGH-based MIN's with the assumption of a 2×2 switch SNR of 20 dB. X-Bar, crossbar.

shows a schematic diagram of the network architecture, and Fig. 11 shows a schematic diagram of the experimental setup. The switch consists of three cascaded BCGH switch arrays, two PR's, and four photodetectors. The fan-out (splitting) stage of the network is an array of 1×2 switches, and it consists of BCGH1 and a PR. The second stage of the network is an array of 2×2 switches. It is constructed by use of BCGH2 and BCGH3 together with a PR. PC of the beams occurs on four photodetectors (one for each output channel) that serve as the output stage of the network. In general, passive optical fan-in is feasible only for smaller networks (see Table 2). The SNR can be increased by use of a stage of active fan-in modules. Note that the 4×4 implementation used a butterflytype interconnect, instead of the shuffle. This permitted smaller deflection angles and allowed a single type of element to be used when the network was folded into two dimensions.

The three BCGH arrays were identical four-level phase, polarization-selective diffractive elements, as shown schematically in Fig. 12. The BCGH switching elements were fabricated by use of standard microfabrication technologies: electron-beam lithography was used to define the mask patterns; optical lithography was employed to transfer the pattern onto the Y-cut lithium niobate substrate; the surface-relief profile was obtained through the use of ion-beam etching. Each BCGH array consisted of a 4×4 array of pixels, where each pixel corresponds to a 1×2 switch. The dimensions of each pixel were approximately $4 \text{ mm} \times 4 \text{ mm}$, so the overall element had an active area of $1.6 \text{ cm} \times 1.6 \text{ cm}$ (Fig. 13).

Each of the 16 switches in an array was a diffractive polarization beam splitter designed to propagate vertically polarized light (solid lines in Fig. 11) straight and to deflect horizontally polarized light (dashed lines) at an angle. The grating period was $40 \text{ }\mu\text{m}$, and the operating wavelength was 514.5 nm . These parameters set the optimum distance between the two BCGH arrays to be approximately 320 mm . Figure 14 shows a photograph of the system that

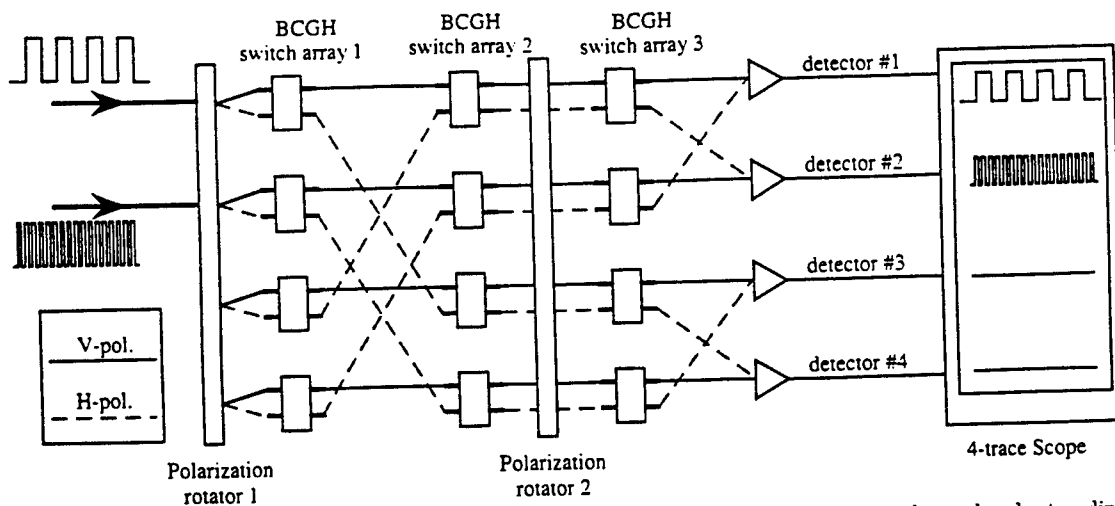


Fig. 10. Schematic diagram of a 4×4 BCGH Stretch switch. The demonstration system in this study employed a two-dimensional folded version of the optical 2-shuffle interconnection. Scope, oscilloscope; V-pol., vertically polarized; H-pol., horizontally polarized.

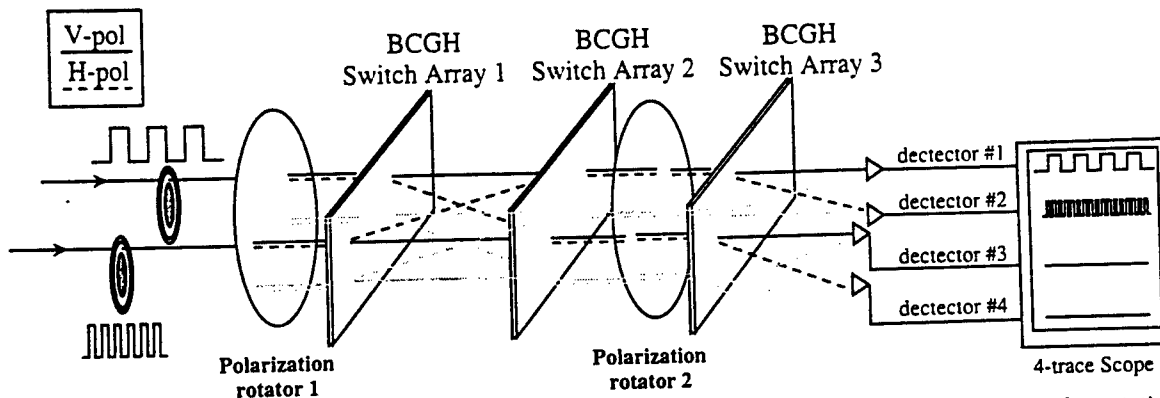


Fig. 11. Diagram of the optical system for the 4×4 switch implementation for tracing two input paths. In the implementation, the network was folded into a two-dimensional array with equal spacing in the horizontal and vertical directions. Scope, oscilloscope; V-pol., vertically polarized; H-pol., horizontally polarized.

displays the cascaded BCGH arrays. A collimated laser beam was used in the experiment, without any relay imaging optics between the BCGH holograms. The measured diffraction efficiencies of the BCGH

holograms were approximately 26%, with peak polarization contrast ratios of 130:1.

Two broadband manual PR's and two electrically controlled liquid-crystal polarization rotators (LCPR's) were used to characterize the switching network and to demonstrate the reconfiguration of the network, respectively. The contrast ratio of the PR's (Newport, Model PR-550) and the Hughes LCPR's were 1000:1 (30 dB) and 4:1 (6 dB), respectively. A beam from an Ar^+ laser was split into two paths, and two mechanical beam choppers modulated at 300 and 900 Hz were used to label the two input beams. To characterize the performance of the network, the SNR was measured at each output of the network by the intensity ratio between the ON state (one of the two input signals is routed to this output node) and the OFF state (both inputs are routed to other output nodes). The worst-case SNR at the output was measured to be 10:1 (10 dB) when both PR's were manual; the best-case SNR was 20:1 (13 dB). During normal

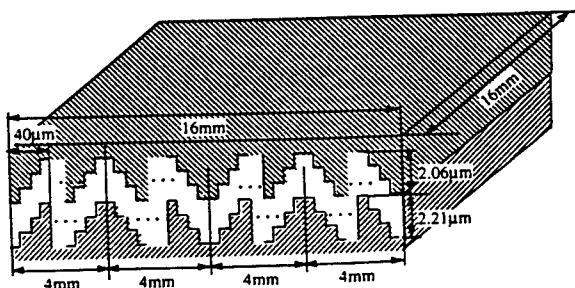


Fig. 12. Schematic cross section of the BCGH used for the MIN studied here. The BCGH was fabricated in LiNbO_3 . The substrate thickness was 1 mm, and the operating wavelength was 514.5 nm.

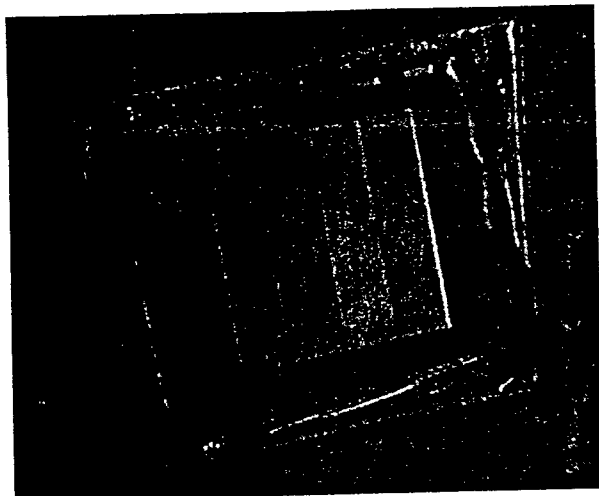


Fig. 13. Photograph of the 4×4 BCGH element array. The dimensions of the array are 16 mm \times 16 mm.

network operation, LCPR's were used to provide electrically controlled reconfiguration. The 4×4 switch was tested with one and two active inputs. Figure 15 demonstrates one active input to the 4×4 switch being switched (reconfigured) between the four outputs. Figure 16 shows the output data with each input beam being switched between two corresponding outputs of the network. In both cases, the worst-case SNR was approximately 4:1, limited by the contrast ratio of the LCPR's.

The reconfiguration speed of the network was also limited by the temporal response of the PR's. The LCPR was operated at a maximum reconfiguration rate of approximately 1 KHz. Using PLZT (lead lanthanum zirconate titanate) or multiple-quantum-well PR's may make reconfiguration as fast as 10–100 MHz possible. Once these PR's were set at a specific configuration, the data rate was limited by

the frequency responses of the transmitters and receivers, because no signal regeneration was used inside the multistage switch. To demonstrate this, we used an acousto-optic modulator modulated at 20 MHz by a pseudorandom non-return-to-zero data generator to modulate one of the input signals. The eye diagram obtained at one of the outputs of the three-stage multistage interconnection network is shown in Fig. 17. Table 3 lists the performance values required of a 2×2 BCGH switch for large switches ($N \geq 1024$) and the best experimental results obtained to date. Note that the increased cross talk of approximately 3 dB for the experimental 4×4 Stretch switch versus that of the 2×2 switch is consistent with the predicted values from Table 2 (AS/PC).

5. Summary

This paper describes the design and implementation of a nonblocking space-division three-dimensional photonic multistage network architecture that uses 2×2 BCGH polarization-selective switches to switch and route the light at each node. The switch architecture uses a fan-out stage, a single stage of 2×2 switches, and a fan-in stage. This architecture is well suited for parallel optical implementation in that (a) it is nonblocking; (b) it enables fast, parallel path hunting with low latency communication; (c) it uses simple 2-shuffle and $N/2$ -shuffle connection patterns (or their equivalents); (d) it uses one stage of $N^2/2$, 2×2 switching elements; and (e) it reduces the effects of first- and second-order cross talk. The fan-out stage may be passive (i.e., simple optical broadcasting), which results in a $2/N$ power efficiency, or can incorporate N fan-out modules (one per input port), where each fan-out module uses a tree-based architecture with $N - 1$, 1×2 switches. Similarly, the fan-in stage can either be active or can use $N/2$ separate detectors per fan-in module. The re-

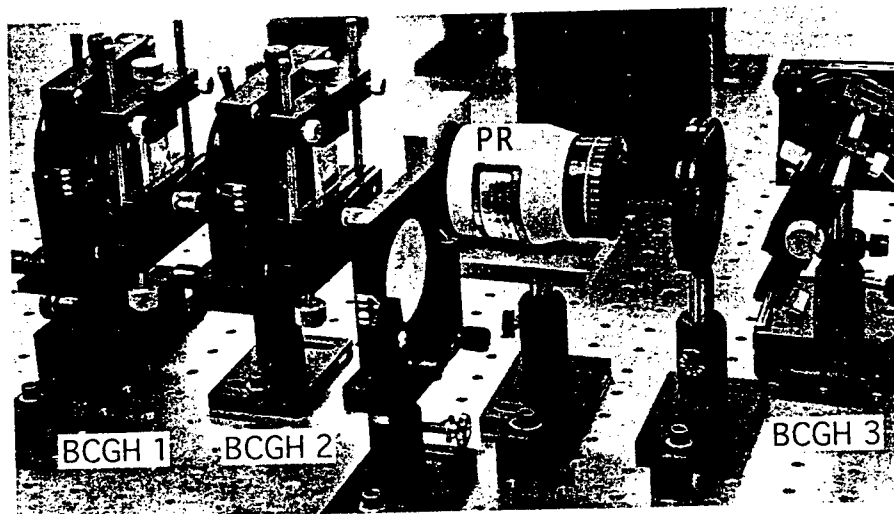
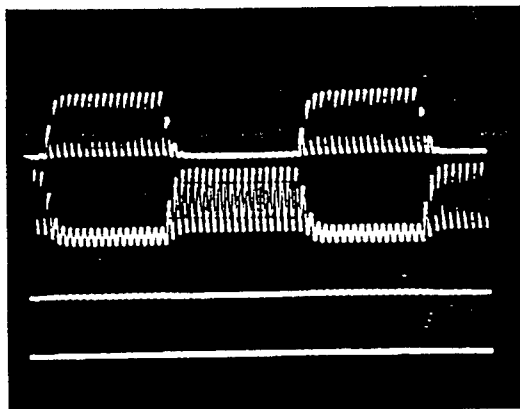
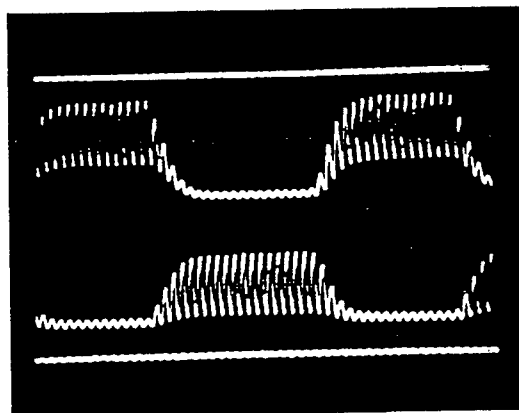


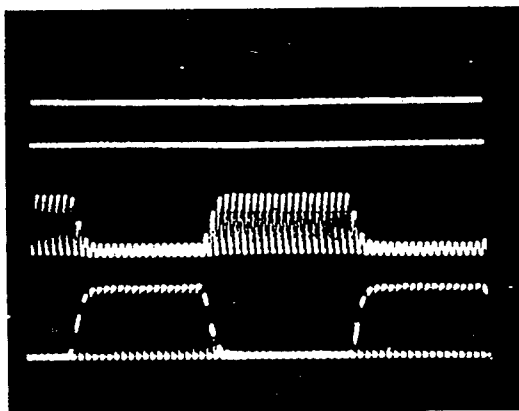
Fig. 14. Photograph of three cascaded BCGH arrays that formed the core of the demonstration free-space switch. The manual PR was used to characterize the network and was replaced by an electrically controlled LCPR for fast reconfiguration.



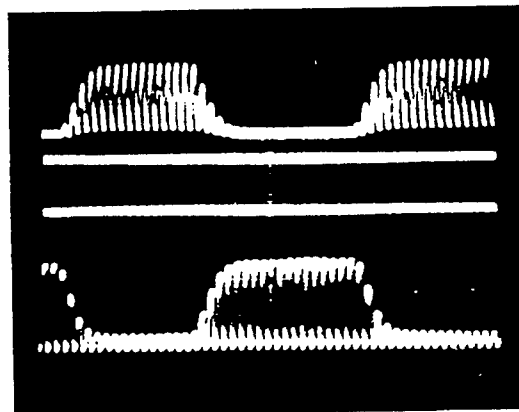
(a)



(b)



(c)



(d)

Fig. 15. Four-trace oscilloscope photographs showing the outputs of the 4×4 switch. An input to the 4×4 switch is being switched (reconfigured) between (a) outputs 1 and 2, (b) outputs 2 and 3, (c) outputs 3 and 4, and (d) outputs 1 and 4. Network cross talk was limited by the 4:1 contrast ratio of the polarization rotator. A SNR of 13 db was measured with manual PR's. The horizontal sweep rate is 10 ms/division.

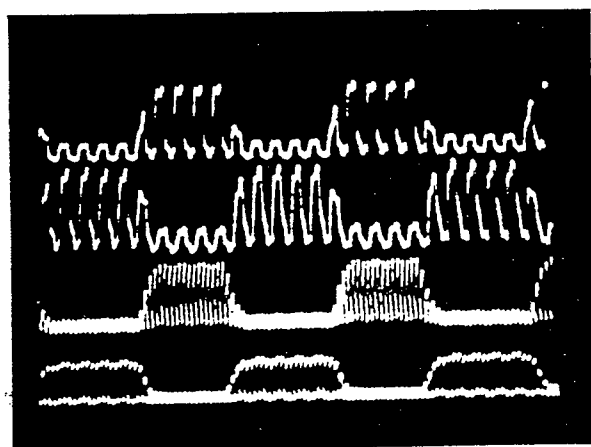


Fig. 16. Four-trace oscilloscope photographs showing two active inputs to the 4×4 switch being simultaneously switched between two outputs of the network. The horizontal sweep rate is 10 ms/division.

sulting photonic switch is circuit switched and can be either locally or externally controlled. The control choice will drive the required light-modulator technology and required pixel complexity. Network setup and reconfiguration times depend on the specific polarization-modulator technology, but after the polarization switches are set, the switching network is transparent, and the data-transmission rate is limited by the source and receiver response.

A small-scale network was demonstrated in an experimental 4×4 BCGH switch. The use of a high-performance pixellated polarization-modulator array, together with ongoing research on improving the performance of the BCGH elements, could make such a switch (with $N \geq 32$ ports) a useful candidate for high-speed optical networks, as well as for large-scale optical multiprocessor interconnection networks.

F. Xu and Y. Fainman acknowledge partial support from Rome Laboratories and the National Science Foundation.

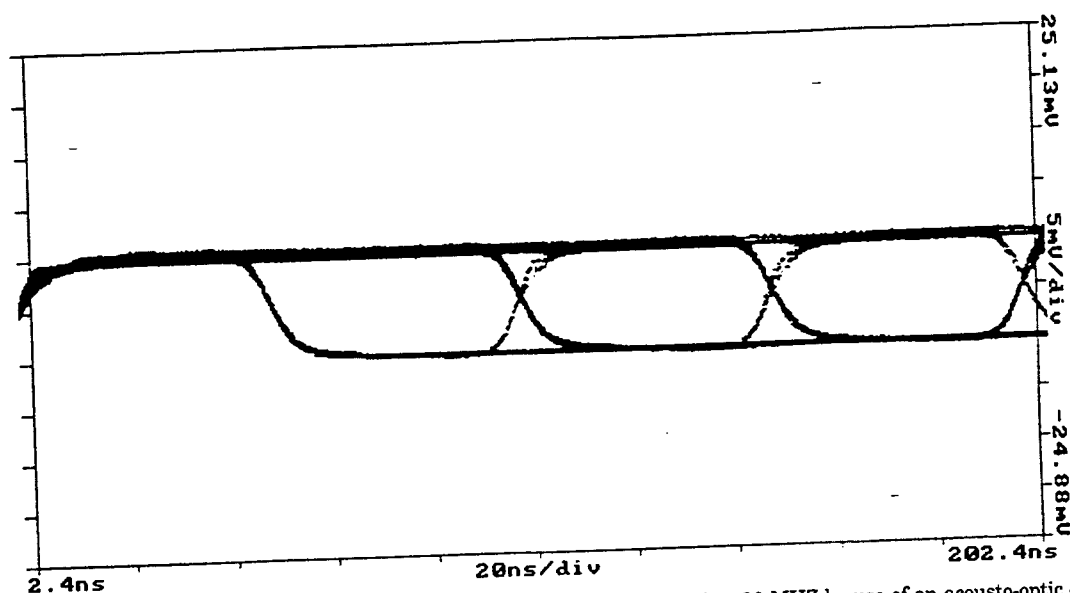


Fig. 17. Eye diagram of a switch output. The input was a laser modulated at 20 MHz by use of an acousto-optic cell.

Table 3. Performance Estimates for a Scalable Switch ($N \geq 1024$ Channels) and Best Experimental Results to Date for 1×2 , 2×2 , and 4×4 BCGH Switches

Performance Estimates	Experimental Results
$\eta = 98\%$	$\eta \approx 60\%$ (four-level phase)
$R_g = R_m = 99\%$	$R_g = R_m \approx 99\%$
$C = 98\%$	$C = 98\%$
2×2 switch efficiency of 98%	2×2 switch efficiency of $\approx 36\%$ insertion loss of -4 dB
insertion loss of -0.64 dB	1×2 switch SNR of 22 dB (160:1)
2×2 switch SNR of 20–30 dB	2×2 switch SNR of 16 dB (40:1)
	4×4 switch SNR of 13 dB (20:1)

References

- See, for instance, IEEE J. Lightwave Technol., vol. 11, nos. 5 and 6, (1993).
- F. Heismann, A. F. Ambrose, To. O. Murphy, and M. S. Whalen, "Polarization independent photonic switching system using fast automatic polarization controllers," IEEE Photon. Technol. Lett. 5, 1341–1343 (1993).
- G. A. De Biase, "Optical multistage interconnection networks for large-scale multiprocessor systems," Appl. Opt. 27, 2017–2021 (1988).
- K. M. Johnson, M. R. Surette, and J. Shamir, "Optical interconnection network using polarization-based ferroelectric liquid crystal gates," Appl. Opt. 27, 1727–1733 (1988).
- T. Nishi, T. Yamamoto, and S. Kuronagi, "A polarization-controlled free-space photonic switch based on a PI-LOSS switch," IEEE Photon. Technol. Lett. 5, (1993).
- K. Noguchi, T. Sakano, and T. Matsumoto, "A rearrangeable multi-channel free-space optical switch based on multistage network configuration," IEEE J. Lightwave Technol. 9, 1726–1732 (1991).
- Y. Wu, L. Liu, and Z. Wang, "Modified gamma network and its optical implementation," Appl. Opt. 32, 7194–7199 (1993).
- T. Sakano, K. Kimura, K. Noguchi, and N. Naito, "256 \times 256 turnover-type free-space multichannel optical switch based on polarization control using liquid-crystal spatial light modulators," Appl. Opt. 34, 2581–2589 (1995).
- R. K. Kostuk, M. Kato, and Y. T. Huang, "Polarization properties of substrate mode holographic interconnects," Appl. Opt. 29, 3848–3854 (1990).
- Y.-T. Huang, "Polarization-selective volume holograms: general design," Appl. Opt. 33, 2115–2120 (1995).
- T. Todorov, L. Nikolava, and N. Tomova, "Polarization holography. 1: A new high-efficiency organic material with reversible photoinduced birefringence," Appl. Opt. 23, 4309–4591 (1984).
- Q. W. Song, M. C. Lee, P. J. Talbot, and E. Tam, "Optical switching with photorefractive polarization holograms," Opt. Lett. 16, 1228–1230 (1991).
- J. Hossfeld, D. Columbus, H. Sprave, T. Tschudi, and W. Dultz, "Polarizing computer-generated holograms," Opt. Eng. 32, 1835–1837 (1993).
- J. E. Ford, F. Xu, K. Urquhart, and Y. Fainman, "Polarization selective computer generated holograms," Opt. Lett. 18, 456–458 (1992).
- J. E. Ford, F. Xu, A. Krishnamoorthy, K. Urquhart, and Y. Fainman, "Polarization-selective computer generated holograms for optical multistage interconnection networks," in Optical Computing, Vol. 7 of OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1993), pp. 258–261.
- F. Xu, J. E. Ford, and Y. Fainman, "Polarization-selective computer-generated holograms: design, fabrication, and applications," Appl. Opt. 34, 256–266 (1995).
- N. Nieuborg, C. Van de Poel, A. Kirk, H. Thienpont, and I. Veretennicoff, "Polarization-selective diffractive and computer-

- generated optical elements," in *Optical Computing*, Vol. 10 of OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1995), pp. 124-126.
18. F. Xu, R. Tyan, P. C. Sun, Y. Fainman, C. Cheng, and A. Scherer, "Form birefringence of periodic dielectric nanostructures," *Opt. Lett.* **20**, 2457-2459 (1995).
 19. G. Bromwell and J. Heath, "Classification categories and historical development of circuit switching topologies," *Comput. Sur.* **15**, (1983).
 20. T. J. Cloonan, G. W. Richards, F. B. McCormick, and A. Lentine, "Architectural considerations for an optical extended generalized shuffle network based on 2-modules," in *Photonic Switching*, H. S. Hinton and J. W. Goodman, eds., Vol. 8 of OSA Proceedings Series (Optical Society of America, Washington, D.C., 1991), pp. 154-157.
 21. A. V. Krishnamoorthy, P. Marchand, F. Kiamilev, and S. Esener, "Grain-size considerations for optoelectronic multistage interconnection networks," *Appl. Opt.* **31**, 5480-5507 (1992).
 22. G. J. Swanson, "Binary optics technology: theory and design of multi-level diffractive elements," DARPA Tech. Rep. 854 (Defense Advanced Research Projects Agency, Washington, D.C., 1989).
 23. L. Bhuyan and D. Agrawal, "Generalized shuffle networks," *IEEE Trans. Comput.* **C-32**, 1081-1090 (1983).
 24. A. V. Krishnamoorthy and F. Kiamilev "Fanout, replication, and buffer-sizing for a class of self-routing packet-switched multistage photonic switch fabrics," in *Photonic Switching*, H. S. Hinton and J. W. Goodman, eds., Vol. 8 of OSA Proceedings Series (Optical Society of America, Washington, D.C., 1991), pp. 87-89. March 1995.
 25. A. V. Krishnamoorthy, "3-dimensional optoelectronic N, M, F networks for neurocomputing and parallel processing," Ph.D. dissertation (University of California, San Diego, San Diego, Calif., 1993).
 26. R. A. Spanke, "Architectures for large non-blocking optical space switches," *IEEE J. Quantum Electron.* **22**, 964-967 (1986).
 27. T. J. Cloonan, F. McCormick, and A. Lentine, "Control injection schemes for photonic switching architectures," in *Photonic Switching*, H. S. Hinton and J. W. Goodman, eds., Vol. 8 of OSA Proceedings Series (Optical Society of America, Washington, D.C., 1991), pp. 162-165.
 28. R. A. Spanke and V. Benes, "N-stage planar optical permutation network," *Appl. Opt.* **27**, 1226-1229 (1987).
 29. V. E. Benes, "Growth, complexity and performance of telephone connecting networks," *Bell Sys. Tech. J.* **62**, 499-539 (1983).
 30. K. Padmanaphan and A. Netraveli, "Dilated networks for photonic switching," *IEEE Trans. Commun.* **30**, 1357-1365 (1987).
 31. S. C. Knauer, A. Huang, and J. H. O'Neill, "Self-routing switching network," in *CMOS VLSI Design*, N. Weste and K. Eshraghian, eds. (Addison-Wesley, Reading, Mass., 1988), Chap. 9, pp. 424-448.
 32. N. K. Ailawadi, "Photonic switching architectures and their comparison," in *Frontiers in Computing Systems Research*, S. Tewksbury, ed. (Plenum, New York, 1990), Vol. 1, pp. 129-186.
 33. A. V. Krishnamoorthy, F. Xu, J. Ford, and Y. Fainman, "Polarization-controlled multistage interconnection network based on birefringent computer generated holograms," in *Photonics for Processors, Neural Networks, and Memories*, J. L. Horner, B. Javidi, S. T. Knowel, and W. J. Miceli, eds., *Proc. SPIE* **2297**, 345-349 (1994).

A. 2. Dynamic Performance of a Circuit-Switched Interconnection Network

P. Dietrich and R. Rao

International Conference on High Performance Computing in New Delhi, India,
December 1995.

DYNAMIC PERFORMANCE OF A CIRCUIT-SWITCHED INTERCONNECTION NETWORK

PAUL DIETRICH
dietrich@cts4.ucsd.edu

RAMESH R. RAO*
rrao@ucsd.edu

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, California, 92093

Abstract

An accurate comparison between blocking and non-blocking interconnection networks requires an evaluation of the additional overhead imposed on the system by the resubmission of blocked requests. By examining a dominant system which bounds the performance of the actual system, we formulate and solve a queuing model for a realistic processor/memory interconnection system using the gated-hold protocol. Additionally, because the solution of this model requires moderate computing power, an approximate solution that agrees closely with the exact analysis and that incorporates non-uniform input and output distributions is described. The approximate solution is validated with simulations for the case of uniform input and output distributions. Results indicate that delay increases approximately as \sqrt{N} instead of $\log(N)$ making this protocol ill suited to large networks.

1 Introduction

Multistage Interconnection Networks (MINs) have been utilized in Parallel Processing Systems to facilitate connection between processors and among processors and memories. A survey of a number of MIN architectures may be found in [6]. Of particular interest are blocking and non-blocking networks. Of the two, non-blocking networks, although more costly to build, offer full connectivity between any free input and output regardless of the traffic pattern. On the other hand, blocking networks are cheaper to build, but, depending on the traffic pattern, messages may be blocked due to contention within the network. It is natural to enhance the functionality of blocking networks to ensure message delivery before comparing them with

non-blocking architectures.

One possible solution is to resubmit blocked messages until eventual delivery. Because requests may have to be submitted several times, each request will experience a *random delay* depending on the prevailing traffic pattern. The precise overhead incurred by this process of resubmissions is not easily determined because of the stochastic nature of the blocking. Dietrich and Rao [11] [12] analyzed a synchronous, circuit-switched square Banyan network of 2×2 crossbar switches implementing a gated-hold protocol in which partial path information is retained to speed circuit set-up. In this protocol, once a set of request enters the network, the network is closed (gated) and no new requests enter until all current requests have been serviced. In [11], the authors derived the *mean* time to fully service a batch of uniform and independent requests appearing at the input to the network. Subsequently, in [12], the authors extend this analysis to incorporate arbitrary independent input distributions and 'hot spot' output distributions.

In a realistic processor-memory interconnect, the number of requests present at the beginning of a cycle is likely to be proportional to the length of the previous cycle. To further extend our understanding of a processor-memory interconnect system, it is imperative that we understand the *temporal evolution* of the request generation process and its effect on the response of the system. This extension, which has not been studied before, is the subject of this paper. In this paper, we model the temporal evolution of the request generation process by assuming that a processor can have only a small number B of memory access requests queued at a time, [7]. We analyze the performance of a circuit switched interconnection network which utilizes the gated-hold protocol and is subject to such an arrival process.

1.1 Proposed Protocol

A synchronous circuit-switched Delta network with a holding protocol was presented in [11] in the context of

*Supported by Air Force Rome Laboratories under contract F30602-95-R0065

a system that allowed for no queuing. This protocol will now be briefly described. Assume that time is divided into periods. At the beginning of each period, the processors submit their requests for connection to the memory devices. As these requests propagate through the multi-stage-interconnection network, some are blocked and others progress. The requests that are blocked hold their partial paths. After the requests that did not get blocked are serviced, the blocked requests continue their advance, starting with those in the stage of the MIN closest to the outputs of the network. This process continues until all requests have been served at which time a new period begins and the processors submit new requests that may have arrived. Unlike the synchronous circuit-switched protocol with request resubmission or dropping, these service periods are not of equal length, and depend upon the number of collisions that occur during contention for the communications paths. Such a service period will be called a network cycle.

1.2 Request Generation and Service Model

We adopt a model in which each processor generates a request with a fixed probability in each fixed length *switching cycle*. Furthermore, at most one request can accumulate over the variable length *network cycle*. Thus, longer the network cycle, larger will be the number of accumulated requests. Therefore, we are able to capture the temporal evolution of the request generation process that one would expect to encounter in a MIN that employs a blocked access scheme. The temporal evolution of the system is as follows.

- In each switching cycle, each processor generates a request that is uniformly destined to any of the memory modules with probability q_i , $1 \leq i \leq N$ independent of other devices, and independent of any requests that are currently in the network. Therefore, it is possible for a processor to generate a new request even while it has a message in the network awaiting transmission. However, once a processor has generated a request in a network cycle, it cannot generate additional requests until the current network cycle ends and the pending request enters the network for transmission. Consequently, at most one memory reference request can be waiting from each processor at the beginning of a service period. In fact, at the end of a service period that required L switching cycles, each processor has a request with probability $1 - (1 - q_i)^L$.
- At the beginning of a network cycle, all waiting requests enter the network and attempt to establish connections. The network uses the gated hold protocol

described earlier. Therefore, all requests that enter in a network cycle will be serviced before its completion. Consequently, each network cycle requires a random number of switching cycles reflecting the number and destinations of the active requests.

- If, at the beginning of a network cycle, no requests are present at the input to the network, we assume that the network waits one switching cycle to allow new requests to be generated and resumes operation. The idle switching cycle does not waste bandwidth nor increase delay since, in this model, no new requests arrive at the input of the network until the instant before the new cycle begins.

We measure the communication delay (D_i) from the instant at which the message is generated by the processor at node i to the instant at which the network cycle in which that requests is served finishes. The delay of a particular message is, in most cases, significantly less than D_i , but this measure provides us with a conservative estimate on communication delay.

The throughput of each input port (T_i) is defined to be the average number of messages that are serviced at node i in $\log(N)$ switching cycles. This aggregate measure will enable us to compare this protocol with other interconnection schemes.

2 Analysis

In principle, an exact Markovian analysis of the proposed system is possible. However, the number of system states that need to be tracked grows exponentially with the number of inputs, hence this approach is not very viable. Consequently, we adopted a different approach. Based on the observation that most of the analytical hurdles follow from the stochastic dependencies within the system, we formulated a worse case "dominating system" that exhibits a reduced degree of dependency.

A simpler structure can be derived if we consider an addition strategy in which the switch, upon seeing a single request at its outputs, produces another request at its other output with a certain probability irrespective of whether or not blocking occurred within the switch.

More specifically, the reduced degree of coupling in the dominating system allows us to derive a recursive equation that expresses the time to service a set of requests at the input to j stages of an N -input network in terms of the time to service a set of requests at the input the $j - 1$ remaining stage of the same N -input network.

2.1 Analytical Technique

Define a random variable $g^N(p, j)$ to be the number of switching cycles required to set up circuits for a set of requests with independent and uniform distributions with marginal input probability p (i.e. the total time required to set up connections for all input devices requiring such connections) at the input to the remaining j stages of a square (N input) Banyan network. A switching cycle is defined as the time taken for a request to propagate through one stage of the switching network. Throughout this analysis, a switching cycle is assumed to take one unit of time.

In [11], a recursive equation for the mean time to service a batch of uniform requests in the dominating system is derived. This result was later extended in [12] to incorporate non-uniform input and output probabilities. Using similar techniques, in [13] a recursive equation for the conditional probability mass function was derived for uniform input probabilities and uniform destination distributions. In the next section, we utilize the result of the analyses summarized in Section 2.1, to study the dynamic evolution of the system as successive batches of requests are generated and held in a queue until they are serviced.

2.2 Queueing Analysis of a Processor/Memory Interconnect

Define the random variable S_n to be the length in switching cycles of the n^{th} network cycle. The length of the cycle is determined by the number of requests present at its beginning. This length in turn, is determined by the length of the previous cycle. Formally, we can show that $\{S_n\}_{n=1}^{\infty}$ is a Markov Chain. The Markov Chain S is aperiodic and irreducible. It has a finite state space and consequently, is ergodic and possesses a stationary distribution π_j . To find the stationary statistics of this chain, we need to compute the transition probabilities P_{ij} . These can be obtained from the recursive solution for the probability mass function presented in [13] for uniform input and routing probabilities only. Although they are a bit tedious to compute, these transition probabilities are easily determined for small networks. When the network size increases the exact computation becomes harder to perform and alternative approaches must be explored. Further details of the exact analysis are omitted for brevity.

2.2.1 Drift Analysis

Because the transition probabilities of the Markov Chain S are difficult to compute for large networks, we consider another technique that yields the approximate steady state

behavior of the chain. The quantity

$$f(i) \triangleq E[S_n - S_{n-1} | S_{n-1} = i]$$

is defined as the drift of the chain S . If this function is concave ($f''(i) \geq 0$), the zero of this function upper bounds the steady state average of the chain S .

To show this, we utilize a result from [3]. Since the chain S is ergodic and has a finite state space, we have from [3]

$$\sum \pi_j f(j) = E_j[E[S_N - S_{n-1} | S_{n-1} = j]] = 0 \quad (1)$$

If $f(i)$ is concave, then by 1 and Jensen's inequality,

$$\begin{aligned} 0 &= E_i[E[S_n - S_{n-1} | S_{n-1} = i]] \\ &\leq E[S_n - S_{n-1} | S_{n-1} = E[S]] \end{aligned} \quad (2)$$

Furthermore, if the function $f(i)$, is such that

$$\begin{aligned} f(i) &\geq 0 \quad \forall j \leq j^* \\ &\leq 0 \quad \forall j \geq j^* \end{aligned}$$

then from (2) we have

$$j^* \geq E[S] \quad (3)$$

The drift can be expressed in terms of the recursive equation presented in [12] as

$$\begin{aligned} f(i) &= E[S_n - S_{n-1} | S_{n-1} = i] \\ &= \sum_{j=2}^{2(N-1)} j Pr[g^N(1 - (1-q)^i, \alpha, k) = j] \\ &\quad + 1 (Pr[g^N(1 - (1-q)^i, \alpha, k) = 1]) \\ &\quad + Pr[g^N(1 - (1-q)^i, \alpha, k) = 0] - i \\ &= E[g^N(1 - (1-q)^i, \alpha, k)] \\ &\quad + \prod_{j=1}^N (1 - q_j)^i - i \end{aligned}$$

where the vector operation are taken elementwise.

Because we do not have a analytic form for the function $f(i)$, it would be a difficult exercise to formally prove the two properties required to show that this is a bound. However, based on graphical evidence we can then show that

$$\begin{aligned} E[W_i] &\geq \frac{j^*}{1 - (1 - q_i)^{j^*}} - \frac{1}{q_i} \\ \text{and} \\ E[T_i] &\simeq \text{Log}(N) \frac{1 - (1 - q_i)^{j^*}}{j^*} \end{aligned}$$

using Jensen's inequality as a bound and approximation respectively.

For comparison, we computed the exact value of the steady state average cycle time and compared it to that found by this approximate bounding techniques for uniform input and output distributions. The closeness of the curves suggest that the approximation, which is significantly easier to compute, provides an excellent measure of steady state system performance.

3 Results

For small networks, the exact system analysis, omitted here for brevity, and the approximate analysis of Section 2.2.1 are very close. Figures 1 and 2 show that the two curves are almost indistinguishable. Because of the difficulties in computing the exact analysis for large networks, we display only those results from the approximate analysis for large values of N in the succeeding figures.

Since these MINs have $\text{Log}(N)$ stages, we can expect the delay to grow with network size at least as fast as $\text{Log}(N)$. Although we have not found an analytic asymptotic expression for $E[g^N(1, k)]$, we can compute $E[g^N(1, k)]$ for extremely large values of N ($N > 2^{30}$). Figure 6 shows these curves to be proportional to $O(\sqrt{N})$. For large values of N , this quantity, hence the total communications delay, grows like $O(\sqrt{N})$ (Fig. 3). It appears that for relatively small values of N , the maximum delay grows like $O(\text{Log}(N)\sqrt{N})$. Additionally, as is evident from Fig. 4, the throughput of the network diminishes as order \sqrt{N} .

The delay/throughput performance of several small size networks is shown in Fig. 5. We see delay increasing as \sqrt{N} and throughput decreasing as \sqrt{N} . As such, this protocol may be poorly suited for use in extremely large interconnection networks. The importance of tracking both the delay and the throughput is evident from this figure. In the absence of such an analysis, one obtains figures for the maximum throughput not recognizing that sometimes operating a little below the maximum can result in significant reduction in communication delay.

4 Conclusions

In this paper, a circuit-switched multistage processor/memory interconnection system was studied under a dynamic request generation model. The temporal evolution of the request generation process was modelled by assuming that the likelihood of a processor generating a request was a function of the length of the previous network cycle. The network utilized a gated-hold protocol in which partial path information was retained to speed circuit set-up time. Approximate and exact bounds for system delay

and throughput were derived. Quantitative results suggest that the delay grows as $O(\sqrt{N})$ when the network size becomes large. The derivation of an asymptotic expression for $E[f^N(1, k)]$ as N grows large would identify asymptotic delay and throughput expressions. However, since the maximum delay and throughput can be computed from $E[f^N(1, k)]$, we can compute delay and throughput bound for all manageable size networks $N \approx 1 \times 10^9$. Poor performance at very large network size limits the use of this protocol to small networks.

The analysis presented in this paper provides a way of computing figures for the maximum throughput and communication delay. Knowledge of the delay throughput curve makes it possible to choose an operating point below the maximum throughput point to reduce communication delay. Furthermore, the bounding methodology used in this paper is evidence of the usefulness of the technique in studying the stochastic evolution of these networks.

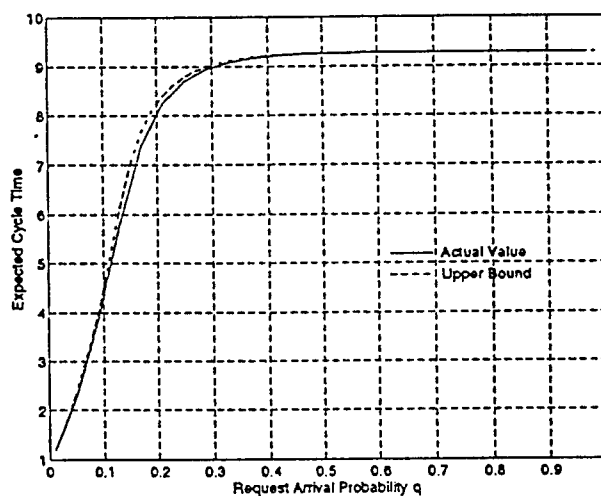


Figure 1: Actual steady state cycle time versus approximate steady state, $N = 8$.

References

- [1] Janak H. Patel, "Performance of Processor-Memory Interconnections for Multiprocessors," *IEEE Trans. on Comput.* Vol C-30, No. 10, Oct. 1981, pp. 771-780.
- [2] Clyde P. Kruskal and Marc Snir, "The Performance of Multistage Interconnection Networks for Multiprocessors," *IEEE Trans. on Comput.*, Vol C-32, NO 12, Dec. 1983, pp. 1091-1098.

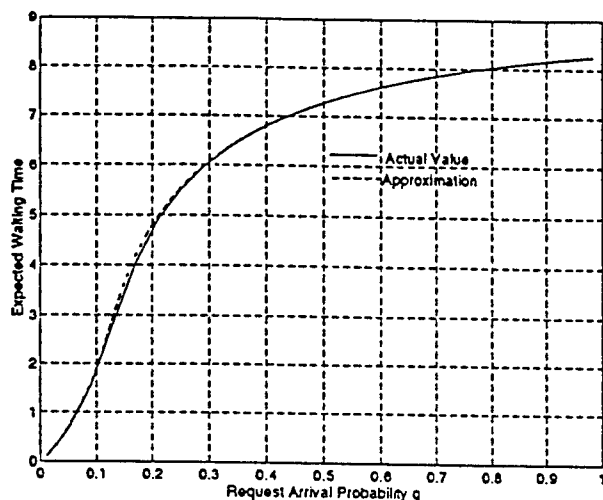


Figure 2: Actual steady state waiting time versus approximate steady state, $N = 8$.

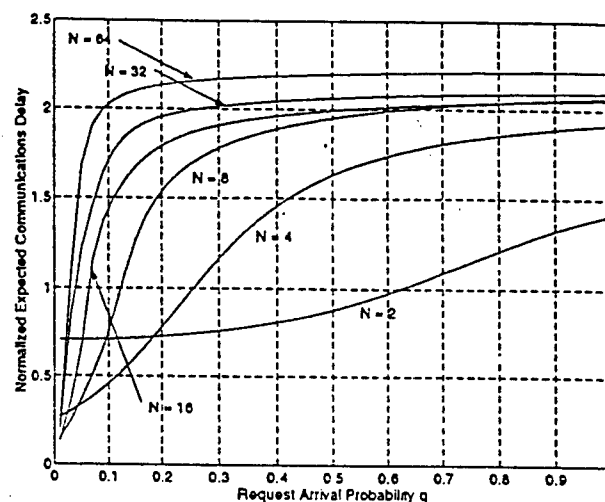


Figure 3: Approximate steady state communications delay normalized by N^{-5} versus request arrival probability q .

- [3] Linn I. Sennott, Pierre A. Humblet, and Richard L. Tweedie, "Mean Drifts and the Non-Ergodicity of Markov Chains," *Operations Research*, Vol. 31, No. 4, July-August 1983, pp. 783 - 789.
- [4] A. Pombortsis and C. Halatsis, "Behaviour of Circuit-Switched Multistage Networks In Presence of Memory Hot-Spot," *Electronic Letters*, Vol. 25, No. 13, June 22, 1989, pp 833-834.
- [5] Peter G. Harrison and Naresh M. Patel, "The Representation of Multistage Interconnection Networks in Queueing Models of Parallel Systems," *Journal of the Association for Computing Machinery*, Vol 37, No. 4, Oct. 1990, pp. 863-898.
- [6] Yuju Oie, Tatsuya Suda, etc., "Survey of Switching Techniques in High-Speed Networks and Their Performance," *International Journal of Satellite Communications*, Vol. 9, 1991, pp. 285-303.
- [7] P. Heidelberger and P.A. Franaszek, "Traffic Studies of Unbuffered Delta Networks," *IBM Journal Of Research and Development*, Vol. 35, No. 1/2, January/March 1991, pp 288-299.
- [8] Chuan-lin Wu and Manjai Lee, "Performance Analysis of Multistage Interconnection Network Configurations and Operations," *IEEE Trans. on Comput.*, Vol-41, No. 1, Jan. 1992, pp. 18-26.
- [9] Shuo-Hsien Hsiao and C. Y. Roger Chen, "Performance Evaluation of Circuit Switched Multistage Interconnection Networks Using A Hold Strategy," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 3, No. 5, Sept. 1992, pp. 632 - 640.
- [10] Clyde P. Kruskal and Marc Snir, "Cost Performance Tradeoffs for Interconnection Networks," *Discrete Applied Mathematics*, 37/38 (1992) pp. 359-385.
- [11] Paul Dietrich and R. R. Rao, "Delay Analysis of a Circuit Switched Interconnection Network Implementing a Gated Hold Strategy," *Allerton Conference on Communications, Computing, and Control*, Allerton, September, 1993.
- [12] Paul Dietrich and R. R. Rao, "Delay Analysis of a Circuit-Switched Interconnection Network with Non-Uniform Traffic," *Proceedings of Infocom '94*, Toronto, Canada, June 12-16, 1994.
- [13] Paul Dietrich and R. R. Rao, "Request Resubmission in a Blocking, Circuit-Switched, Interconnection Network", manuscript under review and available on request.

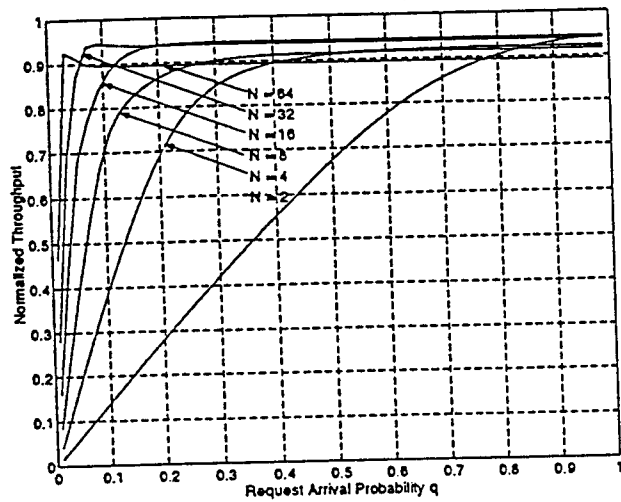


Figure 4: Approximate steady state throughput normalized by N^{-5} versus requests arrival probability q .

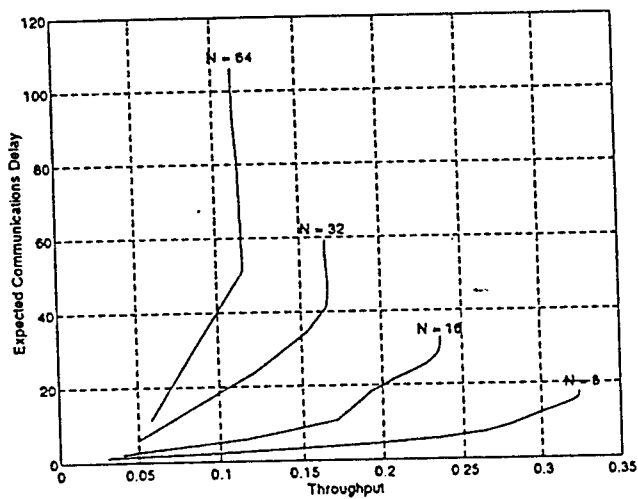


Figure 5: Delay/Throughput characteristics of processor memory interconnect system.

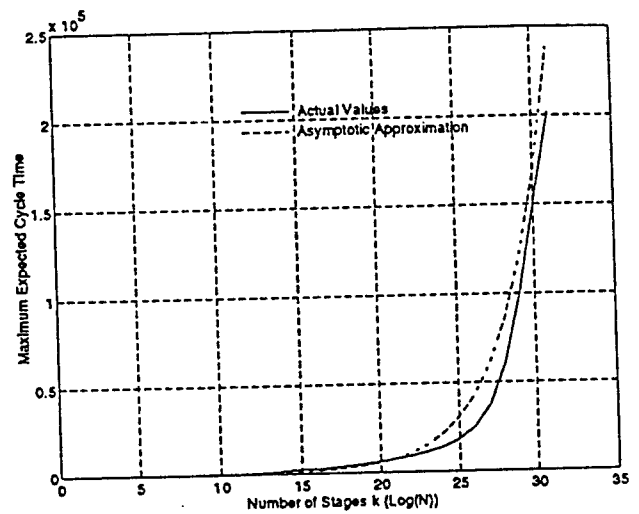


Figure 6: Actual values of $E[f^N(1, k)]$ for large k versus asymptotic approximation.

A. 3. Request Resubmission in a Blocking, Circuit-Switched, Interconnection Network

P. Dietrich and R. Rao

IEEE Transactions on Computers, Vol.45, No.11, November 1996.

Request Resubmission in a Blocking, Circuit-Switched, Interconnection Network

Paul Dietrich and Ramesh R. Rao

Abstract—In this paper, we study the delay performance of a circuit switched, self-routing Delta network in which all requests are guaranteed service. A gated hold protocol, that retains partial path information, is used to guarantee service. A novel technique, that involves the construction of an easier to analyze dominant system, is presented. A recursive expression for the probability mass function of the cycle time in the dominant system is derived. Comparison of the dominant system analysis with simulation of the actual system shows that the dominant system accurately predicts performance for low network loads. As network loads increase, the dominant system becomes worse at predicting behavior of the actual system. These results also help develop insight into how to trade off higher delay variability for increased throughput.

Keywords—Parallel Processing, Resubmission, Non-Blocking, Protocols, Interconnection Network, Performance.

A. Previous Research

In his pioneering analysis, Patel [1] ignored the resubmission of blocked requests and derived an expression for the bandwidth of a Banyan network. Kruskal and Snir's analysis [3], extends Patel's work and provides asymptotic call blocking results, but also assumes that blocked requests are not resubmitted. In a synchronous circuit-switched network with *light* traffic, the resubmission of blocked requests is likely to have little effect on system performance; requests are seldom blocked and therefore do not cause strong temporal traffic correlations at the network input. However, under *heavy* traffic conditions, it is likely that a significant portion of requests will be blocked during each cycle causing time correlations that will affect performance. Under these traffic conditions, frequent blocking and hence resubmissions will occur and the *variability* of the communication delay may increase to unacceptable levels. Because, in parallel processing systems, job delay is often determined by the slowest processor, it is desirable to have an interconnection network with low delay variability. It is therefore imperative that we determine the precise impact of resubmissions on system performance.

In their analysis of an asynchronous circuit switched Banyan network, Wu and Lee [6] considered the effect of resubmission of blocked requests. A self-routing Banyan network was studied under the typical assumption that requests which experience contention are "regenerated" randomly at some later time with a random destination distribution. This result was contrasted with a more realistic blocking model in which requests that experience contention continue to persist in the network (using a drop or hold strategy) until they are serviced. Wu and Lee's results show that the "regenerative" assumption overestimates performance (by as much as 30 percent for the small sized networks tabulated).

Wu and Lee model the position of each request in the network as a Markov chain. A separate chain is needed to describe the state of each request in the network. Naturally, the chains of all the requests are dependent, as blocking causes dependence among the requests. However, Wu and Lee assume independence among the chains, and decouple the chains to simplify the analysis. Because the stochastic dependencies among the requests are ignored, the resulting simplified analysis is likely to be optimistic. One can therefore expect that even their results, which show ignoring blocked requests gives optimistic results, are optimistic.

Bhattacharya, Rao, and Lin [9] recently derived an upper bound on message delay for a synchronous circuit-switched Delta network that accommodates the resubmis-

I. INTRODUCTION

Multistage Interconnection Networks (MINs) have been utilized in Parallel Processing Systems to facilitate connection between processors and among processors and memories. Over the years, a number of MIN architectures have been proposed and analyzed. A representative survey may be found in [2]. In order to select between the many existing MIN architectures, one might strive to characterize all the alternatives in terms of a common set of performance metrics. In this regard, at times, one encounters alternatives that are not easily compared because they differ with respect to their *functionality*.

One such instance occurs when attempting to compare blocking and non-blocking networks. Non-blocking networks, although more costly to build, offer full connectivity between any free input and output regardless of the traffic pattern. On the other hand, blocking networks are less expensive to build, but, depending on the traffic pattern, messages may be blocked due to contention within the network. Clearly we must enhance the functionality of blocking networks to guarantee message delivery before comparing them with non-blocking architectures.

One possible solution is to resubmit blocked messages until eventual delivery. Because requests may have to be submitted several times, each request will experience a *random delay* depending on the traffic pattern. The overhead incurred by this process of resubmissions, which is not easily determined because of the stochastic nature of the blocking, is the object of this study.

Supported by NSF under grant NCR-8904029 and by Air Force Rome Laboratories under contract F30602-95-R0065

sion of blocked requests. To preserve uniformity and independence of the message input distribution they assumed that blocked requests are resubmitted after a random number of cycles and that the resubmitted requests re-randomize their destinations. When several devices require connection through the same switch in the MIN, the re-randomization of the blocked requests may yield performance results that are optimistic as the chance of repeated collisions is lowered.

B. Motivation

Recently, Dietrich and Rao [10] analyzed a synchronous, circuit-switched square Banyan network of 2×2 crossbar switches implementing a gated-hold protocol in which partial information about path setup is retained. In this protocol, once a set of requests enters the network, the network is closed (gated) and no new requests enter until all current requests have been serviced. This protocol, like the hold protocol in [6], allows requests to hold their partial paths until the blocking request completes its service. Unlike the asynchronous hold protocol, the protocol presented in [10] is gated. Gating the network may reduce the network throughput, but guarantees that requests are not blocked by other requests that may arrive at the network many cycles later. In [10], the authors derived a bound for the mean time to fully service a batch of requests appearing at the input to the network. This gated scheme preserves the cycle by cycle FCFS property of the synchronous circuit-switched network, and attempts to reduce the delay variability by preventing the mixing of new and blocked requests. However, the gated-hold strategy achieves this at the expense of network utilization. During the end of the network cycle, only a few requests are left inside the network utilizing resources that could be shared with the requests waiting outside. If one can relax the gating and allow new requests to enter after a fixed length of time, chosen to assure that only a small number of requests remain inside the network when the gate is opened, one can expect an increase in throughput performance at a minimal cost in delay variability.

It is clear that we can gain valuable insight into the performance of such a scheme by examining the probability mass function for the time to completely service a batch of requests under the gated hold protocol introduced in [10]. The probability mass function provides complete information about the delay distribution, and as such, is valuable in determining the effect of prematurely truncating the cycle to increase the network throughput. With this information, the network designer can better control the mixing of traffic and thus the delay variability and FCFS nature of the system.

In this paper, we derive an approximate expression for the probability mass function of the time required to service a batch of independent and uniform requests. This technique also makes it possible to derive simple expressions for higher moments of the delay, which provides a designer with direct information on delay variability.

In the next section, we describe the network protocol,

model, and the underlying assumptions used in the analysis. We construct the dominant system in Section III. A recursive expression for the probability mass function of the cycle time is derived in Section IV-C. In Section V, we present the quantitative results for comparison purposes. Finally, discussion and conclusion sections interpret the significance of these results and suggest ways in which these results can aid in designing more efficient versions of this protocol.

II. PROPOSED PROTOCOL

A. Description

A synchronous circuit-switched Delta network with a holding protocol, briefly presented in [10], is now described in some detail. Assume that time is divided into periods. At the beginning of each period, the processors submit their requests for connection to the memory devices. No new requests may be submitted until all these requests have been served. As these requests propagate through the multi-stage-interconnection network, some are blocked and others progress. The requests that are blocked hold their partial paths. After the requests that did not get blocked are serviced, the blocked requests continue their advance, starting with those in the stage of the MIN closest to the outputs of the network. This process continues until all requests have been served at which time a new period begins and the processors submit new requests that may have arrived. Unlike the synchronous circuit-switched protocol with request resubmission or dropping, these service periods are not of equal length, and depend upon the number of collisions that occur during contention for the communications paths. Such a service period will be called a network cycle.

B. Illustration

As an example, consider the set of active users represented by dots in Fig 1a. After one switching cycle, no collisions occur and all requests advance to the next stage (Fig. 1b). During the second switching cycle, some of the requests require connection through the same switch outputs and only 4 of them progress (Fig. 1c). Of those that progress, 1 is blocked in the next switching cycle and the remaining requests transmit their messages and release their circuits requiring d switching cycles (Fig. 1d). In the first switching cycle following this, the blocked request in the last stage advances and transmits its message (Fig. 1e). In the switching cycle following this, both requests in the second stage advance (Fig. 1f). Finally, in the last switching cycle, these two requests set up a circuit and transmit their messages (not shown). At this point a new network cycle begins with a new input distribution.

C. Implementation

To assess the implementation complexity of this protocol, first note that a user that is not blocked requires no information beyond maintaining synchronization with the network. When a user becomes blocked, it "holds" the circuit and needs to determine when to resume progress. This inform-

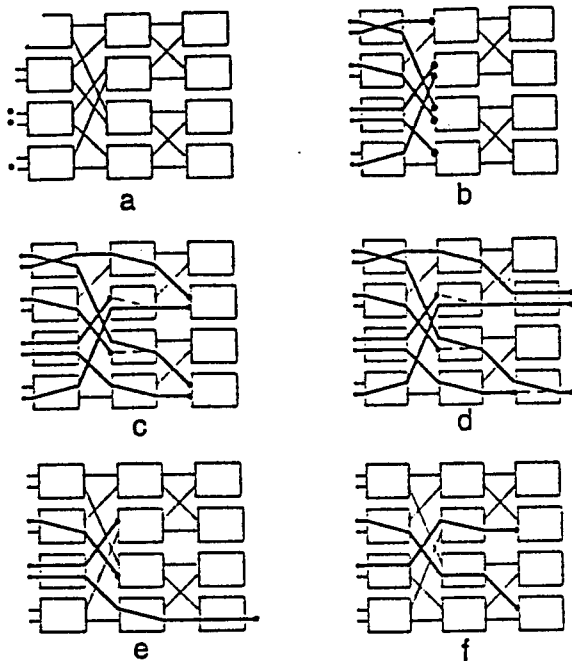


Fig. 1. A Sample path of MIN implementing the gated hold protocol

ation can be derived by maintaining a counter at each stage of the network. The counter must track contentions that may occur at other stages. Additionally, the counter must track whenever a set of requests (or a single request) finish transmission and release their circuits. Because contention can occur in only one stage during any switching cycle, only a single wired-or signal is required to carry contention information to all stages. Similarly, because there will be at most one set of requests that release their circuits during a cycle, only a single signal is required to carry release information.

One implementation of the gated hold protocol is now described. If a request is blocked at a certain stage, it triggers a counter associated with the stage. This counter is initialized to 1. If, during any switching cycles following this event, contention occurs in a stage of switches, the counter is increased by 1. If, during any of the switching cycles, a set of requests release their circuits, the counter is decreased by 1. When the stage counter reaches 0, all requests blocked in that stage are signaled to advance.

The overhead imposed by this protocol is characterized by the hardware complexity required to implement the counters and the wired-or control lines. The up/down counters in an N stage network must count to a maximum of $\log(N)$ requiring a complexity for all the counters of order $\log(N)\log(\log(N))$. The wired or signaling is required at each switch of the network growing in complexity of order $N\log(N)$. Thus, the additional complexity of the network grows at a rate less than or equal to the hardware complexity of the original network.

D. Discussion

To better understand the value of this protocol let us contrast it with protocols that allow dropping. The location and extent of contention within a switch is of course a function of the request generation and destination selection process, with certain request patterns inducing more blocking than others. A switch that is allowed to drop a certain average fraction of its traffic (no matter how small), will operate most efficiently when it selectively drops those patterns that are most contention prone.

Where do these dropped requests go? In a realistic system, blocked requests will be resubmitted at a later time by a "higher layer" entity that recognizes the blocking. This in turn would imply an *accumulation over time of the most difficult to service requests*. The resulting degradation in performance might be quite undesirable. In fact, a study by Heidelberger and Franaszek [5] of a circuit switch revealed that diverting the contention prone traffic to a bypass network resulted in uncharacteristically rapid saturation of the bypass network. Thus, a case can be made that, in the study of protocols that allow dropping, it is necessary to characterize the relative dependencies of the set of dropped requests. Failing this, one has an incomplete understanding of the performance of the interconnection network.

There is another loss of information that occurs when requests are dropped. Typically, when requests collide, the switches allow one group to advance and hold back the remaining. In doing so, a MIN generates the information necessary to partition a group of requests into non-colliding groups via the mechanism of contentions. Dropping partially resolved requests and mixing them with other requests destroys this potentially useful information right after incurring the performance penalty.

Against this backdrop, two key aspects of our protocol stand out. First, by guaranteeing service to all requests we develop an understanding of the *complete* service time. At the same time by retaining path setup and contention information we preserve and exploit available information.

III. CONSTRUCTION OF THE DOMINANT SYSTEM

The difficulty in the analysis of the cycle time is primarily due to the fact that even if requests arriving at a switch are statistically independent, requests departing from a switch are clearly statistically dependent. This is a consequence of the blocking within the switch and implies that the set of requests departing from the first stage are not independent, making the arrival process to the following stages dependent across all switches.

We overcome this hurdle, by constructing a hypothetical a system in which the forwarded and residual requests do not have a complicated interdependency. This makes the hypothetical system more tractable. At the same time, as our results will show, the hypothetical system captures the essential characteristics of the actual system. The basic idea is to find a way to add 'dummy' requests at the outputs of the switches to make the output request distributions *uniform and independent*. If we restrict the strategy

to adding requests and never removing them, then the network cycle time will always be longer. Thus the dominant system would provide an upper bound on delay. Similarly, throughput (of real requests) will be lower for every sample path and the dominant system will provide a lower bound to the system throughput. It must be noted that in the operation of the true system, no 'dummy' requests need be added to the requests.

The dominant system is not designed to remove *all* dependence among requests. Requests within the dominant system still exhibit some dependence. However, the dominant system is designed to force independence among the set of requests at the output of a stage. Similarly, the dominant system is designed to force independence among the set of residual requests at a stage. Jointly, however, the forwarded and residual requests are not independent.

Clearly, adding requests to ensure that at every stage, every input has a request would make requests trivially independent since requests would be present with probability one independent of all other requests. But, such an extreme request addition strategy would not provide reasonable bounds to the performance of the actual system. It is not immediately evident if there exist non-trivial request addition schemes that produces the desired independence for all network loads.

In this section we develop a *non-trivial* request addition strategy that yields useful results for all values of the input probability p . The addition strategy, explicitly specified in Section III-B, requires that each switch, upon seeing a single request at its outputs, adds another request at its empty output with a certain probability irrespective of whether or not blocking occurred within the switch.

This addition strategy was specifically chosen to be independent of whether contention occurred within the switch. An alternative request addition strategy could be based upon using information within the switch regarding collisions. For example, a request could be added on certain ports every time (with probability one) contention occurs. Such request addition strategies were studied and found to be intractable due to the dependence of the strategy on collisions and are not considered in this paper.

A. Notation

- Let X denote a vector of indicator random variables, each of whose elements take the value 1 with probability p independently of others. This vector models the presence of requests at the inputs to the k stage network. The marginal input probability p , which describes the load on the network per input port per cycle, is the main parameter of interest and is explicitly represented in the results.
- Let Y describe the vector of indicator random variables corresponding to those requests that pass out of the first stage after one cycle. Because we have assumed that the inputs to the switches are independent, Y_j will be independent of Y_k for all combinations of outputs k

and j that are not from the same switch. Specifically,

$$\begin{aligned} P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, Y_4 = y_4) \\ = P(Y_1 = y_1, Y_2 = y_2) \cdot P(Y_3 = y_3, Y_4 = y_4) \end{aligned}$$

Each pair from the same switch has the following joint distribution for $i = 1, 2, \dots, N/2$:

$$\begin{aligned} P(Y_{2i-1} = 1, Y_{2i} = 1) &= \frac{1}{2}p^2 \\ P(Y_{2i-1} = 1, Y_{2i} = 0) &= \frac{1}{4}p^2 + p(1-p) \\ P(Y_{2i-1} = 0, Y_{2i} = 1) &= \frac{1}{4}p^2 + p(1-p) \\ P(Y_{2i-1} = 0, Y_{2i} = 0) &= (1-p)^2 \end{aligned}$$

After the requests pass the first stage, the marginal distributions of the requests that advanced remain uniform, but these requests are no longer independent. It is this aspect that we shall focus on next.

B. Injection of Dummy Requests

Consider a dominating system that adds 'dummy' requests at the outputs of the switches to make the output request distributions uniform and independent. Specifically, consider an addition strategy in which the switch, upon seeing a $[1, 0]$ or $[0, 1]$ at its output, independently produces an extra request along the empty path with probability α . The elements of $[top, bottom]$ are indicator random variables signifying the presence or absence of a request at the output of the top and bottom ports of the switch. For example, $[0, 1]$ implies that there is no request present on the top output of the switch but there is a request present on the bottom output of the switch. The destinations of these injected requests are chosen uniformly among all possible destinations.

There are four aspects of the request addition policy that are central to the subsequent analysis.

1. The switch never removes requests, but sometimes adds requests. Consequently, the dominating system can only be worse than the actual system in performance for every sample path.
2. No 'dummy' requests are injected if the switch sees a $[0, 0]$ at its output.
3. Each switch makes its decision regarding the injection of dummy requests independently of the other switches hence, no new dependencies across switches are introduced.
4. Requests additions are made independently of whether or not contention occurred within the switch. Consequently, contention has no effect on the request addition strategy.

The challenge is to find a legitimate probability α such that Y is a vector of uniform, independent random variables. Towards this end, note that the distribution at the switch outputs that results from the injection of new requests in accordance with the strategy just described is

a by:

$$\begin{aligned}
P(Y_{2i-1} = 1, Y_{2i} = 1) &= \frac{p^2}{2} + 2\alpha(p(1-p) + \frac{1}{4}p^2) \\
P(Y_{2i-1} = 1, Y_{2i} = 0) &= (1-\alpha)(p(1-p) + \frac{1}{4}p^2) \\
P(Y_{2i-1} = 0, Y_{2i} = 1) &= (1-\alpha)(p(1-p) + \frac{1}{4}p^2) \\
P(Y_{2i-1} = 0, Y_{2i} = 0) &= (1-p)^2.
\end{aligned}$$

To engineer independence, the joint probabilities must equal the product of the marginal probabilities.

$$\begin{aligned}
P(Y_{2i-1} = 1, Y_{2i} = 1) &= P(Y_{2i-1} = 1)P(Y_{2i} = 1) \\
P(Y_{2i-1} = 0, Y_{2i} = 0) &= P(Y_{2i-1} = 0)P(Y_{2i} = 0).
\end{aligned}$$

The marginal probabilities of Y_{2i-1} and Y_{2i} can be obtained by adding the appropriate joint probabilities. For example, $P(Y_{2i-1} = 1) = P(Y_{2i-1} = 1, Y_{2i} = 1) + P(Y_{2i-1} = 1, Y_{2i} = 0)$. In this way, we can deduce that,

$$P(Y_{2i-k} = 1) = p^2(\frac{3}{4} + \alpha\frac{1}{4}) + p(1-p)(\alpha+1) \quad k = 0, 1. \quad (1)$$

Since the joint probability of both outputs equaling zero is not altered, the marginal probability of equaling zero, $P(Y_j = 0)$, must be the square root of the joint probability or, $(1-p)$. Therefore, $P(Y_j = 0) = (1-p)$. Hence, α we must satisfy the equation:

$$1 - [p^2(\frac{3}{4} + \alpha\frac{1}{4}) + p(1-p)(\alpha+1)] = (1-p). \quad (2)$$

The usefulness of this strategy depends on whether this equation in α has a solution in the interval $(0, 1)$. Solving, we find that $\alpha = p/(4-3p)$. Furthermore, for all values of p in $(0, 1)$, α can be easily verified to be in $(0, 1)$. We have thus established that by using this $\alpha = p/(4-3p)$, we can engineer the distribution at the output of every stage to be uniform and independent. With this choice of α , the marginal probability of having a request remains constant across all stages and equals p .

It turns out that by adding requests to the residual requests in an appropriate manner they too can be made uniform and independent across all inputs. To see this, consider adding to the residual requests as they pass through the first stage of switches (the stage in which they were held back). As with the forwarded requests, the switch adds a request with probability β (β is chosen irrespective of α) if there is a single empty output at the switch. Since there can be a maximum of one request per switch at the inputs, these requests pass through the first stage without contention.

We define a random variable R to be the vector random variable indicating whether a request is present on each output link of the first stage due to the residual requests. Using the request addition strategy described, the pairwise

joint distribution that results is:

$$\begin{aligned}
P(R_{2i-1} = 1, R_{2i} = 1) &= \frac{p^2}{4}2\beta \\
P(R_{2i-1} = 1, R_{2i} = 0) &= (\frac{1}{4}p^2)(1-\beta) \\
P(R_{2i-1} = 0, R_{2i} = 1) &= (\frac{1}{4}p^2)(1-\beta) \\
P(R_{2i-1} = 0, R_{2i} = 0) &= 1 - \frac{p^2}{2}
\end{aligned}$$

To make this joint distribution independent, we must first deduce the marginal probabilities by appropriately summing the joint probabilities. By setting the joint probabilities equal to the product of the marginals we deduce that β must satisfy:

$$\frac{p^2}{4}2\beta + (\frac{1}{4}p^2)(1-\beta) = \sqrt{\frac{p^2}{4}2\beta} \quad (3)$$

For $p \neq 0$, this equation has the solution:

$$\beta = \frac{1 - \sqrt{1 - \frac{p^2}{2}}}{\frac{p^2}{4}} - 1 \quad (4)$$

For $p = 0$ all joint and marginal probabilities are 0 and there is no need to add any 'dummy' requests. It is easily verified for $p \neq 0$, that β is a legitimate probability (i.e. $0 \leq \beta \leq 1$).

By using the β derived above, we can compute the marginal probabilities of the residual requests in the dominant system to be

$$P(R_j = 1) = (1 - \sqrt{1 - \frac{p^2}{2}}). \quad (5)$$

In summary, the fact that α and β are between 0 and 1 for all marginal input probabilities p , proves that a request addition strategy satisfying the four properties stated above exists. In the next section we deduce a recursive expression for the cycle time. In doing so, we will exploit the fact that the forwarded and residual requests are uniform and independent across all inputs and that the respective marginal probabilities are p and $(1 - \sqrt{1 - \frac{p^2}{2}})$.

IV. DERIVATION OF THE DISTRIBUTION FUNCTION

As stated in the introduction, the *probability mass function* provides complete information about the delay distribution, and as such, is valuable in determining a number of performance characteristics including the effect of prematurely truncating the cycle to increase the network throughput. Thus, we need to determine the probability mass function of the time required to service a batch of independent and uniform requests. We shall do so by analyzing the dominant system described in the previous section. The approximation technique presented in this section makes it possible to derive the probability mass function as well as expressions for higher moments of the delay. The results of this section will be compared against simulations in the results section.

A. Notation

Let $f(X, k)$ denote the number of switching cycles required to set up circuits for a set of requests represented by the vector X at the input to k stages of an N -input network. As defined in Section III, each element of X takes the value one with probability p independent of other elements. Our goal is to find the probability mass function of $f(X, k)$. Let the function $F(\cdot)$ denote the set of forwarded requests. Similarly, let $G(\cdot)$ denote the set of residual requests remaining including 'dummy' requests. For convenience, define two indicator random variables $I_F(\cdot)$ and $I_R(\cdot)$, where $I_F(x) = 1$ if the vector x has any non-zero elements and $I_F(x) = 0$ otherwise. $I_R(X) = 1$ if the X requests will result in contention. For convenience we notate the events $\{I_F(\cdot) = 1\}$, $\{I_R(\cdot) = 1\}$, $\{I_F(\cdot) = 0\}$ and $\{I_R(\cdot) = 0\}$, by $I_F(\cdot)$, $I_R(\cdot)$, $I_F(\cdot)^c$ and $I_R(\cdot)^c$ respectively.

B. Recursion

The quantity of primary interest is $Pr[f(X, k) = m]$. Conditioning on the event $\{I_F(X)\}$ we have

$$\begin{aligned} Pr[f(X, k) = m] &= Pr[f(X, k) = m | I_F(X)] \cdot Pr[I_F(X)] \\ &+ Pr[f(X, k) = m | I_F(X)^c] \cdot Pr[I_F(X)^c], \end{aligned} \quad (6)$$

where $Pr[I_F(X)] = 1 - (1 - p)^N$. The conditional probability from the second term of (6) is equal to:

$$Pr[f(X, k) = m | I_F(X)^c] = \begin{cases} 0 & m \neq 0 \\ 1 & m = 0 \end{cases}$$

The conditional probability in the first term of (6) can be further conditioned on $I_R(X)$, defined to be the event that there is blocking as a result of contention in the first set of switches.

$$\begin{aligned} Pr[f(X, k) = m | I_F(X)] &= Pr[f(X, k) = m | I_F(X), I_R(X)] \cdot Pr[I_R(X) | I_F(X)] \\ &+ Pr[f(X, k) = m | I_F(X), I_R(X)^c] \cdot Pr[I_R(X)^c | I_F(X)] \end{aligned} \quad (7)$$

For an initial input distribution with marginal probability p :

$$Pr[I_R(X) | I_F(X)] = \frac{1 - (1 - \frac{p^2}{2})^{\frac{N}{2}}}{1 - (1 - p)^N}. \quad (8)$$

The event $\{I_F(X) = 1\}$ denotes the presence of one or more requests in the vector X . At least one of these requests will be forwarded to the next stage in one switching cycle (there is a winner for every contention). Therefore event $\{I_F(X) = 1\}$ implies that $\{I_F(F(X)) = 1\}$, or written mathematically,

$$\{I_F(X) = 1\} \supseteq \{I_F(F(X)) = 1\}.$$

The second property that we have imposed on the dominant (Sec. III) system guarantees that no requests will be added at a switch if there were no requests at its inputs. This property ensures that the event $\{I_F(F(X)) = 1\}$ implies

that there must have been at least one requests in the vector X or

$$\{I_F(F(X)) = 1\} \supseteq \{I_F(X) = 1\}.$$

Combining these two statements, we can write

$$\{I_F(F(X)) = 1\} = \{I_F(X) = 1\}$$

or in our shorthand notation defined in Section IV-A,

$$I_F(X) = I_F(F(X)).$$

A similar observation can be made about the event $\{I_F(X) = 1, I_R(X) = 1\}$ and $\{I_F(R(X)) = 1\}$. The presence of requests in the vector $F(X)$ along with the event that contention occurred, ensures that there will be at least one requests in vector $R(X)$. Again, because the dominant system adds no requests upon seeing a $[0, 0]$ at its output, the presence of residual requests in the vector $R(X)$ ensures that contention occurred ($\{I_R(X) = 1\}$) and there were original requests in the vector X . Formally, we can write, in the shorthand notation

$$I_F(X), I_R(X) = I_F(R(X)).$$

We can utilize these relationships to further simplify the first term in the expression (7), namely

$$\begin{aligned} &Pr[f(X, k) = m | I_F(X), I_R(X)] \\ &= \begin{cases} Pr[f(X, k) = m | I_F(X), I_F(F(X)), I_R(X)], & 2k \leq m \leq 2(2^k - 1) \\ 0, & \text{otherwise} \\ Pr[f(F(X), k-1) + f(R(X), k-1) & \\ = m - 2[I_R(X), I_F(F(X))], & 2k \leq m \leq 2(2^k - 1) \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (9)$$

In equation (9) the range of m is bounded from below by $2k$ because it is impossible for a network cycle to take less than $2k$ switching cycles given $I_F(X)$ and $I_R(X)$. m is upper bounded by the number of cycles required if there is contention at every switching cycle where contention is possible.

By induction, one can show that the maximum number of cycles required is $2(2^k - 1)$ for a k stage network. To verify the induction, consider a single 2×2 switch. The maximum number of cycles required is 2. Assume that it requires a maximum of $2(2^k - 1)$ cycles for a k stage network. The requests at the input to a $k+1$ stage network will require one cycle to reach the input to the remaining k stages. The residual request left over at the first stage (assuming a worst case) will only require one cycle to reach the input to the remaining k stages. Thus it will require $2 + 2 \cdot 2(2^k - 1)$ or $2(2^{k+1} - 1)$ stages for a $k+1$ stage network, completing the induction.

The last statement of (9) follows, because the residual requests will pass the first stage in one switching cycle since there can be no contention. Conditioned on $I_R(X)$, we

know that there will be both residual and forwarded requests, and therefore, since we have engineered the forwarded and residual request distributions to be uniform and independent given $F(X)$ and $I_R(X)$, the last statement follows.

Equation (9) can be rewritten as a sum over the set $\Gamma \triangleq \{(i, j) : i + j = m - 2\}$ as

$$\begin{aligned}
 & Pr[f(X, k) = m | I_F(X), I_R(X)] \\
 &= \begin{cases} \sum_{(i,j) \in \Gamma} Pr[f(F(X), k-1) = i, \\ f(R(X), k-1) = j | I_F(F(X)), I_R(X)] \\ 2k \leq m \leq 2(2^k - 1) \\ 0, & \text{otherwise} \end{cases} \\
 &\approx \begin{cases} \sum_{(i,j) \in \Gamma} Pr[f(F(X), k-1) = i | I_F(F(X)), I_R(X)] \\ \cdot Pr[f(R(X), k-1) = j | I_F(F(X)), I_R(X)] \\ 2k \leq m \leq 2(2^k - 1) \\ 0, & \text{otherwise} \end{cases} \\
 &= \begin{cases} \sum_{(i,j) \in \Gamma} Pr[f(F(X), k-1) = i | I_F(F(X))] \\ \cdot Pr[f(R(X), k-1) = j | I_F(R(X))] \\ 2k \leq m \leq 2(2^k - 1) \\ 0, & \text{otherwise} \end{cases} \quad (10)
 \end{aligned}$$

The second to last step here is an approximation, since correlations may exist between the forwarded and residual requests. The dominant system does not remove all of the dependence among the requests. We address the accuracy of this approximation in the results section. The last step follows from the engineered independence in the dominant system.

In a similar fashion, we can find the expression for the first term of (7). Noting that even if there is no contention, it will require at least k switching cycles to process the requests given $I_F(X)$, we may deduce that

$$\begin{aligned}
 & Pr[f(X, k) = m | I_F(X), I_R(X)^c] \\
 &= \begin{cases} Pr[f(X, k) = m | I_F(X), I_R(X)^c, I_F(F(X))], \\ k \leq m \leq 2(2^{k-1} - 1) + 1 \\ 0, & \text{otherwise} \end{cases} \\
 &= \begin{cases} Pr[f(X, k) = m | I_F(F(X)), I_R(X)^c], \\ k \leq m \leq 2(2^{k-1} - 1) + 1 \\ 0, & \text{otherwise} \end{cases} \\
 &= \begin{cases} Pr[f(F(X), k-1) = m-1 | I_F(F(X)), I_R(X)^c], \\ k \leq m \leq 2(2^{k-1} - 1) + 1 \\ 0, & \text{otherwise} \end{cases} \\
 &= \begin{cases} Pr[f(F(X), k-1) = m-1 | I_F(F(X))], \\ k \leq m \leq 2(2^{k-1} - 1) + 1 \\ 0, & \text{otherwise} \end{cases} \quad (11)
 \end{aligned}$$

Combining (10) and (11) we have

$$\begin{aligned}
 & Pr[f(X, k) = m | I_F(X)] \\
 &= \begin{cases} 0, & m < k \\ Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \\ \cdot Pr[I_R(X)^c | I_F(X)], & k \leq m < 2k \\ Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \\ \cdot Pr[I_R(X)^c | I_F(X)] \\ + \sum_{(a,b) \in \Gamma} Pr[f(F(X), k-1) = a | I_F(F(X))] \\ \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \\ \cdot Pr[I_R(X) | I_F(F(X))], & 2k \leq m \leq 2(2^{k-1} - 1) + 1 \\ \sum_{(a,b) \in \Gamma} Pr[f(F(X), k-1) = a | I_F(F(X))] \\ \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \\ \cdot Pr[I_R(X) | I_F(F(X))], & 2(2^{k-1} - 1) + 1 \leq m \leq 2(2^k - 1). \end{cases} \quad (12)
 \end{aligned}$$

The probabilities $Pr[I_R(X) | I_F(X)]$ and $Pr[I_R(X)^c | I_F(X)]$ are found explicitly from the marginal probabilities and the independence in the dominant system derived in Section III. Computations of the probabilities follows from first principles. For example, $Pr[I_F(X)]$ is the probability that there is at least one request at the input. Since requests are independent with marginal probability p , this is $(1 - (1-p)^N)$ where N is the number of inputs to the stage. Substituting the expression for $Pr[I_R(X) | I_F(X)]$

and $Pr[I_R(X)^c | I_F(X)]$ derived earlier,

$$Pr[f(X, k) = m | I_F(X)]$$

$$= \begin{cases} 0, & m < k \\ Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \cdot \frac{(1-\frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1-(1-p)^N}, & k \leq m < 2k \\ Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \cdot \frac{(1-\frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1-(1-p)^N} + \sum_{(a,b) \in \Gamma} Pr[f(F(X), k-1) = a | I_F(F(X))] \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \cdot \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N}, & 2k \leq m \leq 2(2^{k-1}-1) + 1 \\ \sum_{(a,b) \in \Gamma} Pr[f(F(X), k-1) = a | I_F(F(X))] \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \cdot \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N}, & 2(2^{k-1}-1) + 1 < m \leq 2(2^k-1). \end{cases} \quad (13)$$

To complete the derivation, we need to compute $Pr[f(X, 1) = m | I_F(X)]$. The total time for all the request to be served through one stage is the maximum (over all switches) of the time required to forward pairs of requests through each switch. Since the inputs to the switches are independent due to the dominant system construction, methods of independent order statistics can be used. Assuming elements of the vector X at the last stage have marginal probability q , we use order statistics to show

$$Pr[f^N(q, 1) = m | I_F(X)] = \begin{cases} 0, & m = 0, \\ \frac{(1-\frac{q^2}{2})^{\frac{N}{2}} - (1-q)^N}{1-(1-q)^N}, & m = 1, \\ \frac{1-(1-\frac{q^2}{2})^{\frac{N}{2}}}{1-(1-q)^N}, & m = 2. \end{cases} \quad (14)$$

Equations (13) and (14) form the recursive solution for the probability mass function of $f(X, k)$ given the event $I_F(X)$. Once this quantity is computed, the conditioning on $I_F(X)$ is removed using (6). This yields an equation for the $Pr[f(X, k) = m]$.

C. Delay Moments in the Dominating System

In this section, we illustrate how the recursive equation (13) just derived can be used to obtain higher moments. We compute the first moment, cross check it against the bound produced in [10], and then compute the second moment.

To find $E[f(X, k)]$ we first determine $E[f(X, k) | I_F(X)]$ and then remove the conditioning. Multiplying by m and

summing on both sides of (13) yields

$$\begin{aligned} E[f(X, k) | I_F(X)] &= \sum_{m=k}^{2k-1} m Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \cdot \frac{(1-\frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1-(1-p)^N} \\ &+ \sum_{m=2k}^{2(2^{k-1}-1)+1} m Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \cdot \frac{(1-\frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1-(1-p)^N} \\ &+ \sum_{m=2k}^{2(2^{k-1}-1)} \sum_{(a,b) \in \Gamma} m Pr[f(F(X), k-1) = a | I_F(F(X))] \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \cdot \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N}. \end{aligned} \quad (15)$$

Combining the first two terms of (15) and substituting s for $m-1$ yields

$$\sum_{s=k-1}^{2(2^{k-1}-1)} (s+1) Pr[f(F(X), k-1) = s | I_F(F(X))] \cdot \frac{(1-\frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1-(1-p)^N}$$

which equals

$$(E[f(F(X), k-1) | I_F(F(X))] + 1) \cdot \frac{(1-\frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1-(1-p)^N}. \quad (16)$$

Examining the last term of (15) it is noted that

$$Pr[f(F(X), k-1) = a | I_F(F(X))]$$

and

$$Pr[f(R(X), k-1) = b | I_F(R(X))]$$

are non-zero only for $k \leq (a, b) \leq 2(2^{k-1}-1)$. Hence

$$\sum_{m=2k}^{2(2^{k-1}-1)} \sum_{(a,b) \in \Gamma} m Pr[f(F(X), k-1) = a | I_F(F(X))] \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \cdot \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N} \quad (17)$$

is equal to

$$\sum_{a=k-1}^{2(2^{k-1}-1)} \sum_{b=k-1}^{2(2^{k-1}-1)} (a+b+2) \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N} \cdot Pr[f(F(X), k-1) = a | I_F(F(X))] \cdot Pr[f(R(X), k-1) = b | I_F(R(X))]$$

which equals

$$(2 + E[f(F(X), k-1) | I_F(F(X))]) \cdot \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N} + E[f(R(X), k-1) | I_F(R(X))] \cdot \frac{1-(1-\frac{p^2}{2})^{\frac{N}{2}}}{1-(1-p)^N}. \quad (19)$$

Combining (16) and (19) we obtain the conditional recursive equation

$$\begin{aligned} E[f(X, k) | I_F(X)] &= (E[f(F(X), k-1) | I_F(F(X))] + 1) \\ &\quad \cdot \frac{(1 - \frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1 - (1-p)^N} \\ &+ (E[f(F(X), k-1) | I_F(F(X))] \\ &\quad + E[f(R(X), k-1) | I_F(R(X))] + 2) \\ &\quad \cdot \frac{1 - (1 - \frac{p^2}{2})^{\frac{N}{2}}}{1 - (1-p)^N} \end{aligned} \quad (20)$$

Multiplying both sides by $1 - (1-p)^N$, removing the conditioning in each term and rearranging terms we get:

$$\begin{aligned} E[f(X, k)] &= E[f(F(X), k-1)] + E[f(R(X), k-1)] \\ &+ 2 - (1 - \frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N. \end{aligned} \quad (21)$$

This equation is identical to the equation for the mean derived in [10]. Thus the approximation (Eqn. 10) made in assuming independence of the forwarded and residual requests does not have any effect on the first moment.

The equation for the second moment and thus the variance of the cycle time can be computed by examining the conditional recursive equations for the first and second moments. The conditional recursive equation for the first moment is shown in (20). To find the recursive equation for the second moment multiply both sides of (13) by m^2 and sum over all values of m . This yields

$$\begin{aligned} E[f(X, k)^2 | I_F(X)] &= \sum_{m=k}^{2k-1} m^2 Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \\ &\quad \cdot \frac{(1 - \frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1 - (1-p)^N} \\ &+ \sum_{m=2k}^{2(2^k-1)+1} m^2 Pr[f(F(X), k-1) = m-1 | I_F(F(X))] \\ &\quad \cdot \frac{(1 - \frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1 - (1-p)^N} \\ &+ \sum_{m=2k}^{2(2^k-1)} \sum_{(a,b) \in \Gamma} m^2 Pr[f(F(X), k-1) = a | I_F(F(X))] \\ &\quad \cdot Pr[f(R(X), k-1) = b | I_F(R(X))] \cdot \frac{1 - (1 - \frac{p^2}{2})^{\frac{N}{2}}}{1 - (1-p)^N}. \end{aligned} \quad (22)$$

Using the same technique used to solve (15) we can solve this yielding

$$\begin{aligned} E[f(X, k)^2 | A_k^p] &= Pr[I_F^c(F(x)) | I_F(X)] \cdot (E[f(F(X), k-1)^2 | I_F(F(X))] \\ &\quad + 2E[f(F(X), k-1) | I_F(F(X))] + 1) \\ &+ (E[f(F(X), k-1)^2 | I_F(F(X))] \end{aligned}$$

$$\begin{aligned} &+ E[f(R(X), k-1)^2 | I_F(R(X))] \\ &+ \{2E[f(F(X), k-1) | I_F(F(X))] \\ &\quad \cdot E[f(R(X), k-1) | I_F(R(X))]\} \\ &+ 4E[f(F(X), k-1) | I_F(F(X))] \\ &\quad + 4E[f(R(X), k-1) | I_F(R(X))] + 4) \\ &\quad \cdot Pr[I_F(F(X)) | I_F(X)]. \end{aligned} \quad (23)$$

To complete the recursion we need to find $E[f(X, 1)^2 | I_F(X)]$. This is computed to be

$$E[f(X, 1)^2 | I_F(X)] = \frac{4 - 3(1 - \frac{p^2}{2})^{\frac{N}{2}} - (1-p)^N}{1 - (1-p)^N} \quad (24)$$

Combining (23) and (20) yields a conditional equation for the variance of the cycle time. The unconditional expressions are then easily derived. In the results section we compare the actual results obtained via simulation with the computed upper bound.

V. RESULTS

We now present the results of the mass function and moment analysis. To examine the tightness of the bounds produced by various techniques, we compare these results to a simulation of the actual gated hold protocol where no additional requests are injected. Simulations were run with 10,000 and 100,000 iterations and showed a maximum difference of 0.6%. Most simulations presented here were run with 10,000 trials.

The mass functions for an 8-input network are shown in Fig 2. A different delay mass function results for each choice of the marginal input probability p . It is clear from this plot, that for moderate values of p , the mass function for the cycle time in the dominant system is *bimodal*. As p increases, and the weight of the mass function shifts to the right, there are situations where the graph shows that two distinct peaks in the mass function exist. To verify the presence of this feature in the delay distribution of the actual system, a simulation of the actual system was conducted.

Results in Fig. 3 4 indicate that the delay distribution in the dominant system and actual system closely match for $N = 32$ and moderate values of p . To more accurately judge the error in the mass function analysis, the distribution function is compared to an estimate of the true distribution function obtained from simulation with 100000 trials. The results show (Fig. 5 and 6) that the distribution function from the analysis of the dominant system bounds the distribution function obtained via simulation. It can also be seen that this bound grows looser with increasing N and p .

The bimodality of the mass functions over certain ranges of p suggests that when the network experiences a moderate amount of contention, requests are served in two bursts. The bimodality suggests that contention is most common at the initial stages resulting in an early splitting into two relatively non-conflicting groups. These two groups could account for the two peaks.

We compare the moments derived from the recursive equation for the mass function of the network cycle time to those obtained by simulation of the actual system. The simulation results are based on 10,000 trials. For the values of N considered, the bounds on the first moment [10] are tight for moderate values of p (Fig. 7). The second moment bounds become looser for increasing N (Fig. 8), but may nonetheless enable system designers to factor in second moment statistics.

Note that although all moments of the dominating system bound the corresponding moments of the actual system, the variance is not a moment and thus it is unclear whether the variance of the dominant system bounds the variance of the actual system. Results clearly show that the variance does not bound the actual system variance. However, the effect of p on the variance of the network cycle time is accurately predicted by the analytical curves. The increase in variance for moderate values of p coincides with the bimodal nature of the mass function in this range. Because the variance is not strictly decreasing in p , there may exist an optimal choice of p in systems that use this protocol. As we suggested in the introduction, systems that must provide a high degree of parallel service often rely upon a set of processors completing their tasks simultaneously. This is easiest to accomplish when the network performance is predictable. Thus if one can tailor the speedup of the network so that it is operating in a low variance mode, better system performance may be obtained.

VI. DISCUSSION

As mentioned in the introduction, this protocol provides cycle by cycle FCFS service and prevents the mixing of requests to reduce delay variability. Some of this may occur at the expense of wasting network resources by holding back requests when only a small number are using the network. If even a small fraction of the old requests are dropped and made to resubmit in the next cycle, the system performance may further improve even though the request distribution will no longer remain independent.

Suppose we were to truncate network cycles, so they never exceed C switching cycles. With a certain probability, upper bounded by the mass function derived in this paper, requests will not be fully served in the C cycles. Assume for simplicity, that these requests are dropped to the beginning of the network for resubmission in the next network cycle. Note that the requests are no longer independent. However, for a low resubmission rate, the dependence will be very small.

To determine the number of requests remaining in the network when the cycle is truncated, consider several possibilities. The simplest bound is to assume that every input port contained a request and that none of the requests were served (this is impossible for $C > k+d$). As a slightly better bound, we could calculate the number of requests that remain assuming all switch inputs had requests with identical destinations. Alternately, as a conservative approximation, one could assume that there are at most Np requests remaining on average. This is the average number submitted

per network cycle which is less than the average number of requests that remain after truncation. We call this an approximation, because, unlike the two sample path bounds above, this is based on expected values and the number of requests remaining at the end of a network cycle is not strictly less than Np .

As an example consider a 16 input network. What is a bound on the dropped request rate when the net cycle is truncated to 9 switching cycles? From the mass function derived earlier, the probability of the network cycle lasting longer than 9 switching cycles is bounded by .04 for a value of $p = .2$. Assuming that we bound the number of requests remaining in the system by the second bound discussed above, there can be at most 12 requests remaining in the system a (This is a very loose bound). Assuming that the dropped requests are equally likely to be from any input port (as the switch does not assign priority), an upper bound on the dropped request resubmission rate is .03 for each port. If we use the conservative approximation, we obtain a dropped request resubmission rate of .008 per port.

It is possible to increase the efficiency of the proposed protocol. To see this note that at times, requests which are held in the network at a particular stage may be able to advance even *before* all requests ahead of them have been serviced. In the above example, two requests are held in the network for several switching cycles when, in fact, they could have advanced to their destinations without contention (Fig. 1e). The protocol described in this paper, however, clearly bounds a protocol incorporating this modification, making it useful in the design of a system using this improvement. Modifying the protocol to allow this sort of stochastic, opportunistic advancement is not pursued in this paper.

VII. CONCLUSIONS

In this paper, we considered a N -input square Delta network implementing a gated-hold protocol and studied the probability mass function of its cycle time. The derivations were validated against simulations.

The quantitative results for $N = 8, 16, 32$, and 64 show that the mass function is, in general, not unimodal. This information, not available from the moments, provides valuable insight into understanding and possibly modifying this protocol to increase its efficiency. The time required to compute the mass function of the delay via this recursive equation grows rapidly with N . However, a technique was demonstrated which allows one to derive recursive equations for arbitrary moments of the delay distribution. The recursive equations for these higher moments require significantly less computational effort.

A bound on the second moment of the gated-hold protocol was explicitly derived here. Although this bound grows looser with increasing N , it provides a measure of comparison between interconnection schemes. It is conceivable that two interconnection networks will have the same average performance, yet when incorporated into a processing system where the performance of the slowest pro-

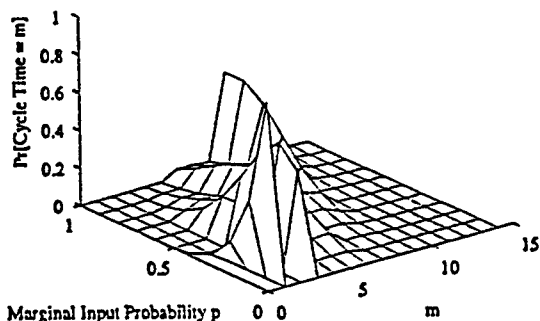


Fig. 2. Network cycle time mass function of an 8-input network versus p ($d = 0$). A delay mass function results for each value of p .

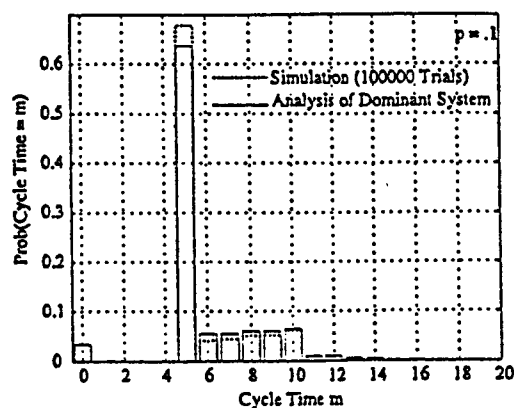


Fig. 3. Network cycle time mass function of an 32-input network for $p = 0.1$ ($d = 0$).

cessor is the limiting design factor (i.e. in the execution of a synchronous algorithm), one network could drastically outperform the other. The second moments statistics can help predict such behavior. Finally, the techniques used in this analysis appear to be quite promising and may find application in other similar systems.

REFERENCES

- [1] Janak H. Patel, "Performance of Processor-Memory Interconnections for Multiprocessors," *IEEE Trans. on Comput.* Vol. C-30, No. 10, Oct. 1981, pp. 771-780.
- [2] Feng, Tse-yun, "A Survey of Interconnection Networks," *Computer*, December 1981, pp. 12 - 28.
- [3] Clyde P. Kruskal and Marc Snir, "The Performance of Multistage Interconnection Networks for Multiprocessors," *IEEE Trans. on Comput.*, Vol. C-32, No. 12, Dec. 1983, pp. 1091-1098.
- [4] Peter G. Harrison and Naresh M. Patel, "The Representation of Multistage Interconnection Networks in Queueing Models of Parallel Systems," *Journal of the Association for Computing Machinery*, Vol. 37, No. 4, Oct. 1990, pp. 863-898.
- [5] P. Heidelberger and P.A. Franaszek, "Traffic Studies of Unbuffered Delta Networks," *IBM Journal Of Research and Development*, Vol. 35, No. 1/2, January/March 1991, pp 288-299.
- [6] Chuan-lin Wu and Manjai Lee, "Performance Analysis of Multistage Interconnection Network Configurations and Operations," *IEEE Trans. on Comput.*, Vol. 41, No. 1, Jan. 1992, pp. 18-26.
- [7] Shuo-Hsien Hsiao and C. Y. Roger Chen, "Performance Evaluation of Circuit Switched Multistage Interconnection Networks Using A Hold Strategy," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 3, No. 5, Sept. 1992, pp. 632 - 640.
- [8] Clyde P. Kruskal and Marc Snir, "Cost Performance Tradeoffs for Interconnection Networks," *Discrete Applied Mathematics*, 37/38 (1992) pp. 359-385.
- [9] Amiya Bhattacharya, Ramesh R. Rao, and Ting-Ting Lin, "Delay Analysis of Synchronous Circuit Switched Delta Networks," *In Proceedings, 1992 International Parallel Processing Symposium*, Newport Beach, Ca., 1992.
- [10] Paul Dietrich and R. R. Rao, "Delay Analysis of a Circuit Switched Interconnection Network Implementing a Gated Hold Strategy," *Allerton Conference on Communications, Computing, and Control*, Allerton, September, 1993.

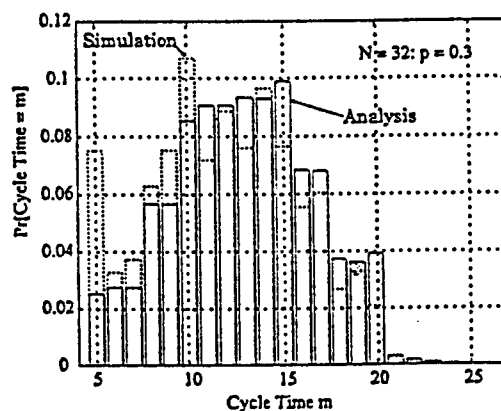


Fig. 4. Network cycle time Distribution function of an 32-input network for $p = 0.3$ ($d = 0$). Comparison of actual versus dominant systems.

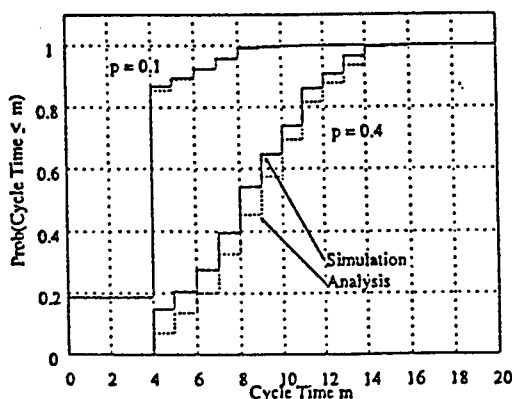


Fig. 5. Network cycle time Distribution function of a 16-input network for $p = 0.1, 0.4$ ($d = 0$). Comparison of actual versus dominant systems.

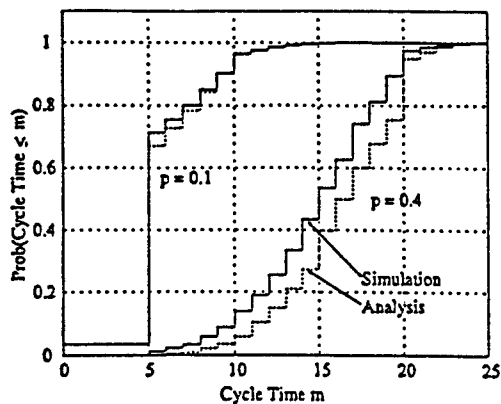


Fig. 6. Network cycle time Distribution function of a 32-input network for $p = 0.1, 0.4$ ($d = 0$). Comparison of actual versus dominant systems.

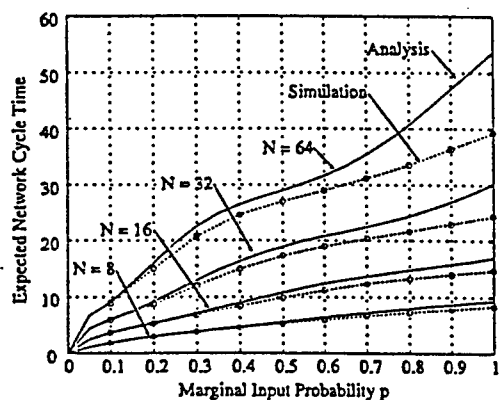


Fig. 7. Expected Network cycle time versus p ($d = 0$). Comparison of actual versus dominant systems for $N = 8, 16, 32, 64$.

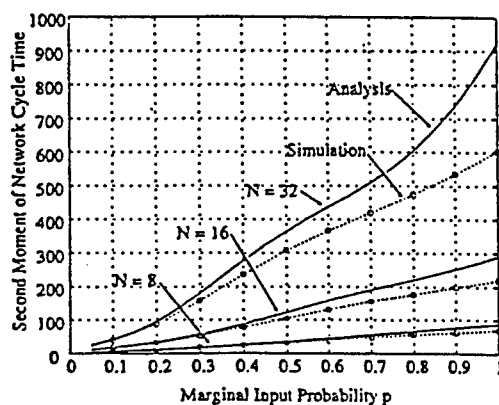


Fig. 8. Comparison of analysis (solid) and simulation (dashed) of second moment of network cycle time for N -input networks ($N = 8, 16, 32$; $d = 0$).

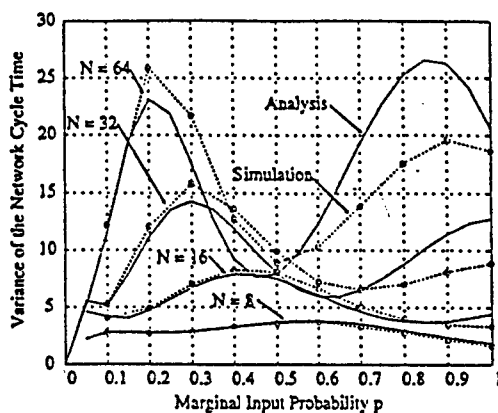
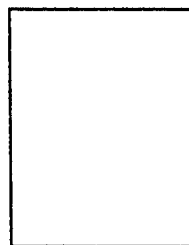
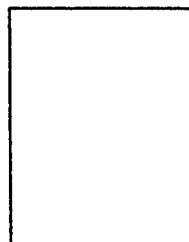


Fig. 9. Comparison of analysis (solid) and simulation (dashed) of variance of the network cycle time for N -input network ($N = 8, 16, 32, 64$; $d = 0$).



stochastic dependence, networks for parallel processing, packet communications over time varying channels, and wireless MAC layer protocols.

†



include architectures, protocols and performance analysis of computer and communication networks.

A. 4. Single Substrate Birefringent Computer Generated Holograms

F. Xu, R.C. Tyan, J. Ford and Y. Fainman

Optics Letters, Vol. 21, No. 7, April 1, 1996.

Multiple Order Delay Holograms for Polarization and Color Selectivity

Fang Xu, Rong-Chung Tyan, Joseph E. Ford*, and Yeshayahu Fainman
Department of Electrical and Computer Engineering
University of California, San Diego, La Jolla, CA 92093-0497

*AT&T Bell Laboratories
101 Crawfords Corner Rd., Holmdel, NJ 07730

Diffractive optical elements constructed as phase only computer generated holograms (CGHs) are attractive for numerous applications in photonics and optoelectronics. A conventional diffractive optical element (DOE) has a maximum phase delay of 2π between pixels. Therefore, the required etch depth is, in general, shallow ($< \text{wavelength } \lambda$, see Fig. 1a). These DOEs are relatively insensitive to the polarization and wavelength of the reconstruction field compared to volume gratings. Previously we demonstrated polarization selective diffractive optical elements using two birefringent LiNbO_3 substrates with different diffractive microstructures on the interior.^{1,2} The required etch depths on both substrates are deeper than that in a conventional DOE because the substrates birefringence is relatively small (see Fig. 1b). Another approach to achieve the same functionality is based on deep etch structure (see Fig. 1c) that corresponds to multiple periods of phase delays (also called modular $2m\pi$) using a single birefringent substrate, as first proposed in reference 1 and later in reference 3. This approach may reduce the cost and simplify the fabrication process of such polarization selective diffractive optical elements. In the following we report the design, fabrication and characterization of multiple order delay (MOD) holograms that possess dual functionality in polarization or color.

To design a MOD hologram with dual impulse responses using single substrate, we use the geometrical optics approximation and find the corresponding phase delays caused by an etched pixel compared to that of an unetched pixel

$$\begin{aligned} k_1(n_1 - n_{11})d &= 2l\pi + \Phi_1 \\ k_2(n_2 - n_{12})d &= 2m\pi + \Phi_2 \end{aligned} \quad (1)$$

where $\Phi_1 + 2l\pi$ and $\Phi_2 + 2m\pi$ are the phase delays exhibited by the two independent optical reconstruction fields, d is etch depth of the pixel, k_1 and k_2 are the wavevectors of the two reconstruction fields, n_1 and n_2 are the refractive indices of the substrate for the two reconstruction fields, n_{11} and n_{12} are the refractive indices of the material surrounding the microstructure, and l and m are integers corresponding to the multiple periods of phase delays exhibited by the two fields. The two independent reconstruction fields can be of different wavelengths or of orthogonal linear polarizations. In general, Eqs. (1) does not have a unique accurate solution for d , if Φ_1 and Φ_2 are arbitrarily specified design values, unless the refractive indexes n_1 and n_2 can be controlled in every pixel of the diffractive element as in a form birefringent artificial dielectric nanostructures⁴. However, for our design with a homogeneous substrate characterized by constant values of n_1 and n_2 , there exist only an approximate solutions for d when the values of integers l and m are arbitrarily large such that

$$\begin{aligned} k_1(n_1 - 1)d &= 2l\pi + \Phi_1 + \delta_1 \\ k_2(n_2 - 1)d &= 2m\pi + \Phi_2 + \delta_2 \end{aligned} \quad (2)$$

where δ_1 and δ_2 are small numbers representing the approximation errors and air is used for the medium surrounding the microstructure thus $n_{12}=n_{11}=1$. If δ_1 and δ_2 are much smaller than the value of the phase quantization level, this is a valid approximate solution to Eqs. (1). With this

approach, it is possible to design the two independent phase functions within some specified accuracy at each pixel. Therefore, independent multilevel phase holograms can be implemented for the two orthogonal polarizations or two different wavelengths. Solving Eq.(2) we find,

$$d = \frac{(2l\pi + \Phi_1 + \delta_1)\lambda_1}{2\pi(n_1 - 1)} = \frac{(2m\pi + \Phi_2 + \delta_2)\lambda_2}{2\pi(n_2 - 1)} \quad (3)$$

We used the above algorithm to design and demonstrate two types of MOD holograms with dual functionality. The first one is a polarization selective element made of a single birefringent yttrium orthovanadate (YVO₄) substrate and the second one is a wavelength selective element made of BK7. YVO₄ has large birefringence and can be relatively easy to process using microfabrication techniques. We used x-cut YVO₄ crystals grown by CASIX, Inc. The $n_o=2.0241$ and $n_e=2.2600$ of YVO₄ were determined at $\lambda=0.5145 \mu\text{m}$. Using Eqs. (2) and (3) with these values of refractive indices, we find all the possible combinations of Φ_1 and Φ_2 that are necessary for construction of a binary phase single substrate birefringent computer generated hologram (SSBCGH) (see Table 1). In Table 1, d_l and d_m are the calculated exact etch depths from Eq. 1 that are required to obtain Φ_1 and Φ_2 for various integers l and m . We observe from Table 1 that by choosing a single value for each etch depth will introduce approximation errors of less than about 6% for all cases. This error can be further reduced by taking the value d as the weighted average of d_l and d_m instead of one half of the summation. Other optimizations such as choosing a different set of phase quantization bases may also reduce the approximation errors. Furthermore, because the absolute phase in diffractive optics is of no concern, we can remove an etch depth bias of $1.013 \mu\text{m}$ (see Table 1) without affecting the desired relative phase values between different pixels. Therefore, only two distinct etches are needed to construct a binary phase level SSBCGH (s and t in Table 1).

For experimental demonstration and characterizations of such a SSBCGH, we constructed a diffractive polarization beam splitter that diffracts one polarization while transmitting the other. This is a special case of the dual functionality element that requires a single value of etch depth. The desired diffractive structure was defined and transferred by electron beam and photolithographic processes, and then the element was ion beam etched to $1.032 \mu\text{m}$. The duty ratio of the fabricated SSBCGH element was measured to be 1:1. The experimental evaluation of the element shows 70.8% diffraction efficiency and 79.7:1 polarization contrast ratio (PCR) into the zero order, 37.4% diffraction efficiency and 33:1 PCR into the +1st order and, 38.9% diffraction efficiency and 32.5:1 PCR into the -1st order.

To better understand the fabrication accuracy requirements and their effect on the performance of the fabricated SSBCGH elements we used rigorous coupled wave analysis (RCWA)⁵ to simulate the performances of our fabricated element. Fig. 2 shows the simulation results for diffraction efficiencies and PCRs as functions of etch depth for grating with 1:1 duty ratio and vertical side-walls. From the simulation results, we can observe that, the good performance (>40% diffraction efficiency and over 100:1 PCR) can be obtained, which is close to our geometrical optics design. The experimental performance of the fabricated element is very close to that predicted by the RCWA (see Fig. 2). From the simulation, we can also see that the performance of a SSBCGH can be further improved with more accurate etch depth. Also, the RCWA results show that the etch depths for the best PCR and the largest diffraction efficiency are very similar but not identical. This important result implies that the desired etch depth can be driven by the application needs and may slightly differ from the values provided by the geometrical optics approximate design listed in Table 1.

Using the same approach, we also demonstrated a wavelength selective element for operation as a color selective beamsplitter for wavelengths of 1.30 μm and 1.55 μm . The substrate material is BK7 glass. The indices of refraction of BK7 were specified by the supplier (Newport Optical Materials Inc.) to be 1.5027 at 1.30 μm , and 1.5004 at 1.55 μm . Using Eq. 1, the phase delay for a 7.75 μm deep etch is 5.994π at 1.3 μm and 5.004π at 1.55 μm . This set of values provides the necessary phase delay for a simple wavelength beam splitter that transmits 1.3 μm light field and deflects 1.55 μm light field. The element was etched to 7.9 μm using chemically assisted ion beam etching method with CHF_3 as the reactive gas.

Figure 3 shows the measured diffraction efficiency and the location of each of the first four orders for the fabricated element. A perfectly fabricated binary phase hologram, neglecting Fresnel reflection losses, should have no energy in the even orders, 40.5% in the ± 1 st orders, and 4.5% in the ± 3 rd orders. At 1.55 μm , the diffraction efficiencies of the fabricated element matched these numbers closely, with 39% in each of the first orders, 3.6% in each of the ± 3 rd orders, and a zero order transmission of 0.83%. At 1.3 μm , the transmission was 83%, while the diffraction into any of the orders was less than 1.2%.

A MOD hologram is more sensitive to changes in the illumination angle than a conventional DOE because of the increased optical path differences. We tested the effect of tilting the element, and found that the performance (first order diffraction efficiency at 1.55 μm and zero order transmission at 1.3 μm) changed by less than 2% for a 5° tilt, and less than 10% for a 10° tilt. In fact, the overall performance was slightly improved with a 5° tilt, suggesting that the etch depth should be increased to 7.93 μm to optimize performance. A field angle of 10° indicates that these elements are compatible with F/3 and larger optics.

In conclusions, we have demonstrated multiple order delay holograms with dual impulse responses in polarization or color. The experimental results indicate good performances. Such elements may be useful in image processing, optoelectronic packaging and photonic switching.

Authors thank P.C. Sun and K. Urquhart for helpful discussions. The research conducted at UCSD is funded by National Science Foundation, Air Force Rome Laboratory and AFOSR.

References:

- 1 J. Ford, F. Xu, K. Urquhart, and Y. Fainman, Opt. Letts., 18:456, (1993).
- 2 F. Xu, J. Ford, and Y. Fainman, Appl. Opt., 34:256, (1995).
- 3 S. Liu and Y. Chen, Opt. Lett., 20:1832, (1995).
- 4 I. Richter, P. Sun, F. Xu, and Y. Fainman, Appl. Opt., 34:2921, (1995).
- 5 M. G. Moharam, T. K. Gaylord, JOSA, 72:1385, (1982).

Φ_o, Φ_e	l, m	d_l, d_m	$d = (d_l + d_m)/2$	$d_r = d - 1.013$	error (%)	
		(μm)	(μm)	(μm)	δ/Φ_o^*	δ_m/Φ_e^*
0, 0	4, 5	2.010, 2.042	2.0260	1.0130 (= s)	+3.27	-3.9
0, π	2, 2	1.005, 1.021	1.0130	0.0000	+1.63	-3.8
π , 0	2, 3	1.256, 1.225	1.2406	0.2276 (= t)	-6.1	+3.8
π , π	4, 5	2.261, 2.246	2.2535	1.2405 (= s+t)	-2.9	+3.8

Table 1 Design results and the real etch depth required for a binary phase single substrate BCGH. (* When Φ_o and/or Φ_e are zero, they are taken to be 2π for errors evaluations.)

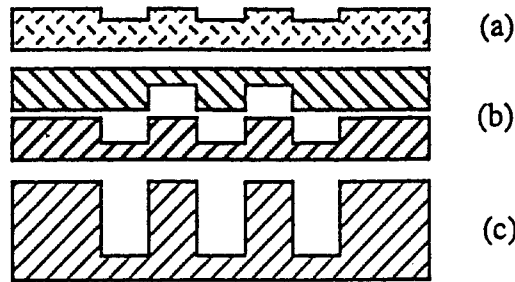


Figure 1. (a) conventional DOE, (b) two-substrate BCGH and (c) multiple order delay DOE.

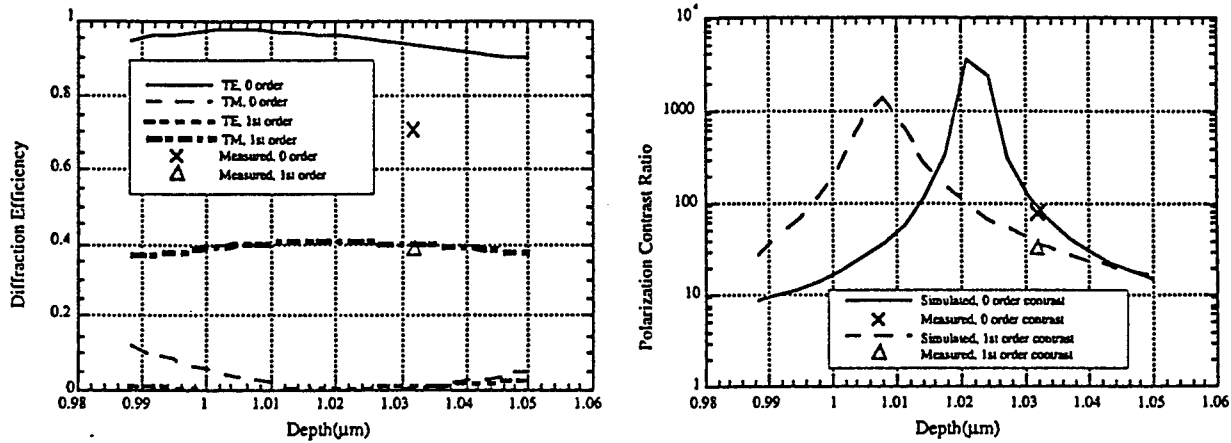


Figure 2. Simulated (curves) and measured (data points) performances of the SSBCGH

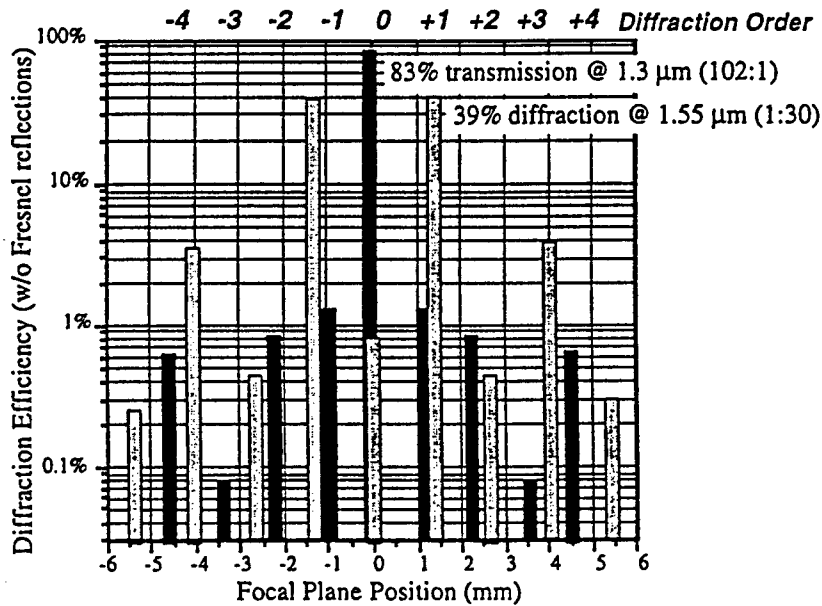


Figure 3. Intensity at diffraction peaks for 1.3 μm (solid) and 1.55 μm (shaded) illumination.

A. 5. Multiple Order Delay Holograms for Polarization and Color Selectivity

F. Xu, R.C. Tyan, J. Ford and Y. Fainman

Presented at OSA Topical Meeting on Diffractive Optics, 1996

Single-substrate birefringent computer-generated holograms

Fang Xu, Rong-Chung Tyan, and Yeshayahu Fainman

Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92093-0407

Joseph E. Ford

AT&T Laboratories, Holmdel, New Jersey 07733

Received October 2, 1995

A polarization-selective computer-generated hologram fabricated upon a single substrate of birefringent YVO₄ crystal is demonstrated. Rigorous couple-wave analysis was used to model the element. The experimentally measured first diffraction order showed a close-to-theoretically predicted diffraction efficiency of 39%. The polarization contrast ratio was measured to be 33:1. © 1996 Optical Society of America

Diffraction optical elements (DOE's) constructed as phase-only computer-generated holograms are attractive for numerous applications in photonics and optoelectronics. In general, planar DOE's are insensitive to the polarization of the illumination. However, polarization-selective computer-generated holograms are attractive for numerous applications because they use polarization as another degree of freedom to implement two independent and arbitrary impulse responses for the two orthogonal linear polarizations. Previously we demonstrated such polarization-selective DOE's, using two birefringent LiNbO₃ substrates.^{1,2} These birefringent computer-generated holographic elements have been used for such applications as transparent photonic switching and networking, image processing, and packaging of optical and optoelectronic devices and systems.¹⁻³

A normal DOE has a maximum phase delay of 2π between pixels. Therefore the required etch depth is, in general, shallow [see Fig. 1(a)]. Previously demonstrated birefringent computer-generated holographic elements consist of two substrates with different diffractive microstructures on the interior, in which the two independent surface-relief depths provide us with the two degrees of freedom to encode the two independent phase functions for the two orthogonal linear polarizations in the same element.^{1,2} The required etch depths on both substrates are deeper than those in a normal DOE because the birefringence is relatively small [see Fig. 1(b)]. Another approach to achieve the same functionality is based on multiple periods of phase delays (also called modular $2m\pi$) with a single birefringent substrate, as first proposed in Ref. 1 and later in Ref. 4. Such an approach may reduce the cost and simplify the fabrication process of such polarization-selective diffractive optical elements. The required etch depth is much deeper than that in a normal DOE [see Fig. 1(c)]. In this Letter we report what to our knowledge is the first experimental demonstration of a polarization-selective computer-generated hologram using a single birefringent substrate. We also describe the design principles, the fabrication procedures, and the experimental characterization results of the single-substrate birefringent computer-generated hologram (SSBCGH).

To design a SSBCGH element, we use a multiorder phase microstructure, in which each pixel of the microstructure is deeply etched such that propagating optical waves will exhibit multiple periods of phase delays. Consider a surface-relief microstructure fabricated in a birefringent substrate with the optic axis parallel to the surface of the substrate. Using geometrical optics, we can find the corresponding phase delays caused by the surface relief compared with those of an unetched pixel for the ordinary- and extraordinary-polarized waves:

$$\begin{aligned} (2\pi/\lambda)(n_o - n_t)d &= 2l\pi + \Phi'_o, \\ (2\pi/\lambda)(n_e - n_t)d &= 2m\pi + \Phi'_e, \end{aligned} \quad (1)$$

where $\Phi'_o + 2l\pi$ and $\Phi'_e + 2m\pi$ are the phase delays exhibited by ordinary and extraordinary waves, λ is the wavelength of the incident wave in vacuum, n_o and n_e are the refractive indices for ordinary- and extraordinary-polarized light, respectively, n_t is the refractive index of the material surrounding the microstructure, and l and m are integers corresponding to the multiple periods of phase delays exhibited by the ordinary- and extraordinary-polarized light. In general, Eqs. (1) do not have unique accurate solutions for d if Φ'_o and Φ'_e are arbitrarily specified de-

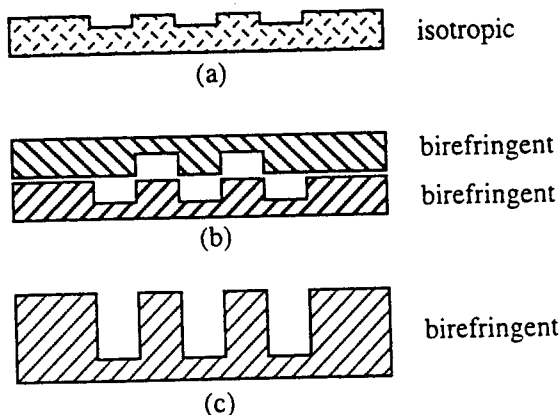


Fig. 1. Schematics of (a) a conventional DOE and (b) two-substrate and (c) a single-substrate birefringent computer-generated hologram.

sign values, unless the refractive indices n_o and n_e can be controlled in every pixel of the diffractive element, as in form-birefringent artificial dielectric nanostructures.⁵ However, for our design with a homogeneous birefringent substrate with constant values of n_o and n_e , there exist approximate solutions for d when the values of integers l and m are arbitrarily large such that

$$(2\pi/\lambda)(n_o - 1)d = 2l\pi + \Phi_o + \delta_l, \\ (2\pi/\lambda)(n_e - 1)d = 2m\pi + \Phi_e + \delta_{1m}, \quad (2)$$

where δ_l and δ_m are small numbers introduced to account for the errors and air is used as the material surrounding the microstructure, $n_i = 1$. If δ_l and δ_m are much smaller than the phase quantization step, we consider this to be a valid approximate solution to Eqs. (1). With this approach it is possible to apply an arbitrary phase with some specified accuracy to each pixel at each polarization. Therefore independent multilevel phase holograms can be implemented for the two polarizations. Solving Eqs. (2), we obtain

$$d = \frac{(2l\pi + \Phi_o + \delta_l)\lambda}{2\pi(n_o - 1)} = \frac{(2m\pi + \Phi_e + \delta_m)\lambda}{2\pi(n_e - 1)}. \quad (3)$$

The ideal substrate material suitable for this application should have large birefringence, to ensure that the required etch depth d can be fabricated with high accuracy. We used a new type of birefringent crystal, yttrium orthovanadate (YVO₄), which has large birefringence and can be relatively easily processed with microfabrication techniques. The birefringent substrates are x-cut YVO₄ grown by Castech-Phoenix, Inc. The refractive indices are $n_o = 2.0241$ and $n_e = 2.2600$ at a wavelength of $0.5145 \mu\text{m}$. Using Eqs. (2) and (3) with these values of refractive indices, we find all the possible combinations of Φ_o and Φ_e that are necessary for construction of a binary-phase-level SSBCGH (see Table 1). In Table 1, d_l and d_m are the exact etch depths required for finding the desired phase delays with integers l and m . We observe that the approximation errors are less than 5% for all cases except one, which we can solve by taking the value d as the weighted average of d_l and d_m instead of one half the summation. Some other optimizations, such as choosing a different set of phase quantization bases, may also reduce the approximation errors. Because the absolute phase in diffractive optics is of no concern, we can remove an etch-depth bias of $1.013 \mu\text{m}$ without affecting the desired relative phase values between different pixels. The new real values of the etch depths d_r are also listed in Table 1. We

can also observe that one needs only two distinct etches to construct a binary-phase-level SSBCGH (s and t in Table 1).

Construction of more efficient multiple-phase-level elements is also possible. Now the minimum etch-depth increments are determined by phase combinations $\Phi_o = 0$, $\Phi_e = 2\pi/N$ and $\Phi_o = 2\pi/N$, $\Phi_e = 0$, where N is the number of phase quantization levels. The required etch depths for other values of phase levels are the combination of these basic ones. However, in some cases these values may result in unrealistically deep etch-depth requirements; moreover, the number of required etches is more than that for a conventional multiple-phase-level DOE.

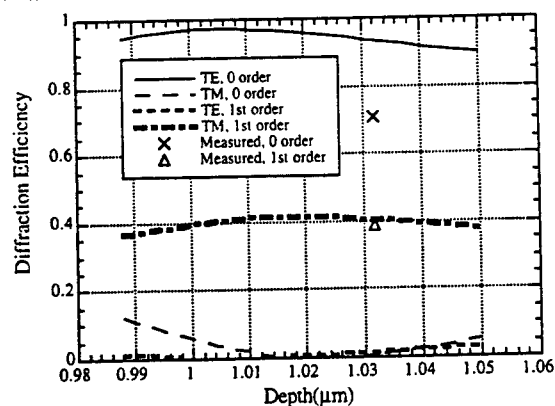
To understand better the fabrication accuracy requirements and their effect of the performance of the fabricated SSBCGH elements, we used rigorous coupled-wave analysis⁶ to simulate the performances of our first fabricated SSBCGH element. Figure 2 shows the simulation results for diffraction efficiencies and polarization contrast ratios (PCR's) as functions of etch depth for a grating with a 1:1 duty ratio and vertical sidewalls. From the simulation results we can observe that good performance ($\sim 41\%$ diffraction efficiency and $>100:1$ PCR) can be obtained from the geometrical optics design. Also, the results of the rigorous coupled-wave analysis show that the etch depths for the best PCR and the largest diffraction efficiency are very similar but not identical ($\sim 1\%$ difference). This important result implies that the desired etch depth can be driven by the application needs and may differ slightly from the values provided by the geometrical optics approximate design listed in Table 1. Additional simulation and experimental evaluations show that the trapezoidal shape and the uneven duty cycle degrade the performance of the SSBCGH significantly.

Figure 3 shows the PCR as a function of the optical wave incidence angle for different grating-period-to-wavelength ratios, calculated with the rigorous coupled-wave analysis. Grating grooves are perpendicular to the incident plane. From this figure we can observe that multiple-order phase-delay elements are sensitive to the incidence angle. These curves also indicate that this design approach is valid only for large grating-period-to-wavelength ratios because, when the grating period is comparable with or smaller than the wavelength of the incident optical field, the form-birefringence effect becomes dominant.⁷ Within this region, not only do the PCR's drop considerably but, in addition, most of the incident wave energy propagates into the zeroth diffraction order.

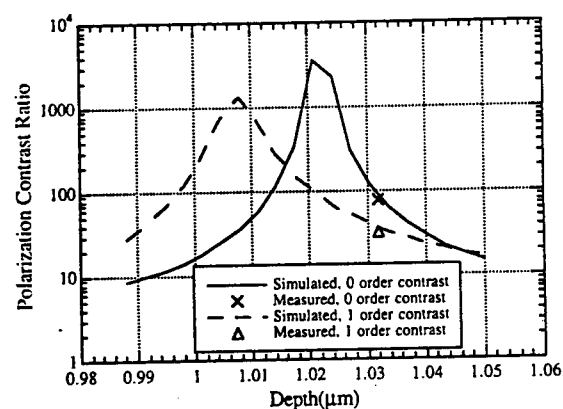
Table 1. Design Results and the Real Etch Depths Required for a Binary-Phase SSBCGH with YVO₄

Φ_o, Φ_e	l, m	$d_l, d_m (\mu\text{m})$	$d = (d_l + d_m)/2 (\mu\text{m})$	$d_r = d - 1.013 (\mu\text{m})$	Error (%) ^a	
					δ_l/Φ_o	δ_m/Φ_e
0, 0	4, 5	2.010, 2.042	2.0260	1.0130 (=s)	+3.27	-3.9
0, π	2, 2	1.005, 1.021	1.0130	0.0000	+1.63	-3.8
π , 0	2, 3	1.256, 1.225	1.2406	0.2276 (=t)	-6.1	+3.8
π , π	4, 5	2.261, 2.246	2.2535	1.2405 (=s + t)	-2.9	+3.8

^aWhen Φ_o and Φ_e are zero, they are taken to be 2π for error evaluations.



(a)



(b)

Fig. 2. (a) Diffraction efficiencies and (b) PCR's of the SSBCGH.

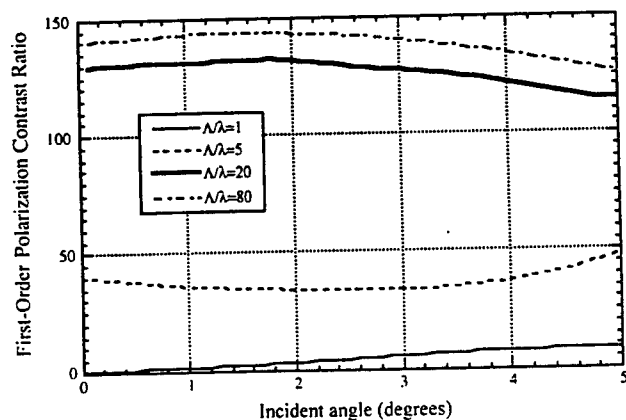


Fig. 3. Simulated PCR's as functions of the incident angle and the grating-period-to-wavelength ratio.

For experimental demonstration and characterization of the SSBCGH, we constructed a diffractive polarization beam splitter that diffracts one polarization while transmitting the other. This is a special case of the dual-function element. First, a

40- μm -period grating pattern defined by electron-beam lithography was photolithographically transferred into a 1.7- μm -thick photoresist. The resist was spun coated onto a 100-nm-thick Cr layer evaporated onto the YVO_4 substrate. Then the surface-relief profile was ion milled into the YVO_4 substrate to 1.032 μm . In this multiple-phase-period approach the etch depth must be controlled with higher accuracy than that used for construction of conventional DOE's. This is because the error introduced by the over/under etch is determined not by the ratio of etch error to the total etch depth but by the ratio of the etch error to a fraction of the total etch depth that is, in effect, responsible for encoding the desired phase values in a given pixel. The duty ratio of the fabricated SSBCGH element was measured to be 1:1. The experimental evaluations of the element show 70.8% diffraction efficiency and 79.7:1 PCR into the zero order, 37.4% diffraction efficiency and 33.0:1 PCR into the +1 order, and 38.9% diffraction efficiency and 32.5:1 PCR into the -1 order. The experimental performance of the element is very close to that predicted by the rigorous coupled-wave analysis (see Fig. 2). From the simulation we can also see that the performance of a SSBCGH can be further improved with more-accurate etch depths.

In conclusion, we designed, fabricated, and experimentally evaluated polarization-selective computer-generated holograms, using a single birefringent substrate. The fabricated elements show diffraction efficiencies close to the theoretical limit and large polarization contrast ratios. The duty ratio and the shape of the grating change the performance of the SSBCGH and need to be controlled accurately. Such elements may be useful in many applications such as image processing, transparent photonic switching, and the packaging of optoelectronic devices and systems.

The authors thank K. Urquhart and P. C. Sun for helpful discussions. This research is funded by the National Science Foundation and the U.S. Air Force Rome Laboratory.

References

1. J. Ford, F. Xu, K. Urquhart, and Y. Fainman, *Opt. Lett.* **18**, 456 (1993).
2. F. Xu, J. Ford, and Y. Fainman, *Appl. Opt.* **34**, 256 (1995).
3. A. Krishnamoorthy, F. Xu, J. Ford, and Y. Fainman, *Proc. SPIE* **2297**, 345 (1994).
4. S. Liu and Y. Chen, *Opt. Lett.* **20**, 1832 (1995).
5. I. Richter, P. Sun, F. Xu, and Y. Fainman, *Appl. Opt.* **34**, 2921 (1995).
6. M. G. Moharam and T. K. Gaylord, *J. Opt. Soc. Am.* **72**, 2921 (1995).
7. F. Xu, R. C. Tyan, P. C. Sun, Y. Fainman, C. C. Cheng, and A. Scherer, *Opt. Lett.* **20**, 2457 (1995).

A. 6. Form-birefringent computer-generated holograms

F. Xu, R.C. Tyan, P.C. Sun, C.C. Cheng, A. Scherer and Y. Fainman

Optics Letters, Vol. 21, No. 18, February 15, 1996

Form-birefringent computer-generated holograms

Fang Xu, Rong-Chung Tyan, Pang-Chen Sun, and Yeshayahu Fainman

Department of Electrical and Computer Engineering, 0407, University of California, San Diego, La Jolla, California 92093-0407

Chuan-Cheng Cheng and Axel Scherer

Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125

Received March 25, 1996

Polarization-selective computer-generated holograms made with form-birefringent nanostructures were designed, fabricated, and evaluated experimentally at 1.5 μm . The fabricated element showed a large polarization contrast ratio ($>250:1$) and a high diffraction efficiency ($>40\%$ for a binary phase level element). The experimental evaluation was in good agreement with the design and modeling predictions. © 1996 Optical Society of America

Polarization-selective phase-only birefringent computer-generated holograms (BCGH's) are general-purpose diffractive elements that have independent impulse responses for orthogonal linear polarizations. Such elements are shown to be useful in many applications, including packing optoelectronic devices or systems, free-space optical interconnects, and image processing.¹ BCGH's have been demonstrated with two birefringent substrates^{2,3} and with a single birefringent substrate.⁴ The birefringence of the substrates in these configurations makes the elements sensitive to the polarization of the light. The two-substrate approach is complicated to fabricate because it includes an assembly process of the two diffractive structures that requires high alignment accuracy. The single-substrate approach, on the other hand, is simpler to fabricate, but it is only an approximate solution. In this Letter we report a new approach for design and fabrication of BCGH elements. Our new approach involves creating a form-birefringent nanostructure and modulation of the refractive index as well as the birefringence at each pixel of the BCGH.

Form birefringence is a well-known effect of subwavelength periodic microstructures. The electric fields parallel to the grating grooves (TE polarization) and perpendicular to the grating grooves (TM polarization) need to satisfy different boundary conditions, resulting in different effective refractive indices for TE- and TM-polarized waves.⁵ Many researchers have demonstrated this effect in the far-IR region.^{6,7} Recently, with the help of the advances in nanofabrication, 200-nm period gratings were fabricated in a GaAs substrate that showed strong form birefringence in the near IR.⁸ Furthermore, these results were found to be in agreement with the numerical simulation results obtained by a rigorous coupled-wave analysis (RCWA).^{9,10} Design optimizations were performed for BCGH by form birefringence.⁹ Chen and Craighead demonstrated a polarization-insensitive diffractive optical element that uses two-dimensional subwavelength periodic microstructures.¹¹ Aoyama and Yamashita demonstrated a grating beam splitting polarizer that uses a subwavelength grating fabricated in a photoresist.¹² The polarization contrast ratios,

defined as the ratio of intensities obtained under two orthogonal polarizations at the designed diffraction order, were $\sim 6:1$ and $\sim 3:1$ for the 0th and 1st diffraction orders, respectively.

In what follows, we report the design, fabrication, and experimental evaluation of a binary phase level BCGH element that uses form-birefringent nanostructures [or form-birefringent computer-generated holograms (FBCGH's)] fabricated upon GaAs substrates for operation in the near-IR wavelength range. The FBCGH element is designed to transmit the TE polarization straight ahead and deflect the TM polarization at an angle.

Consider a single period in a binary phase diffractive structure as shown in Fig. 1. In this period T , one pixel consists of a high-spatial-frequency grating (HSFG) with period Λ , and the other pixel is the substrate material. Of the two periodic structures, the HSFG does not introduce propagating diffraction orders other than the 0th order because of its subwavelength nature. The diffractive structure, on the other hand, introduces many diffraction orders. The phase differences between rays 1 and 2 for TE and TM polarizations are

$$\begin{aligned} (2\pi/\lambda)(n_s - n_{\text{TE}})d &= \Phi_{\text{TE}}, \\ (2\pi/\lambda)(n_s - n_{\text{TM}})d &= \Phi_{\text{TM}}, \end{aligned} \quad (1)$$

where λ is the wavelength in vacuum, d is the thickness of the HSFG layer, n_s is the refractive index of the substrate, and n_{TE} and n_{TM} are the effective refractive indices of the HSFG for TE and TM polarization, respectively. When the wavelength is much larger

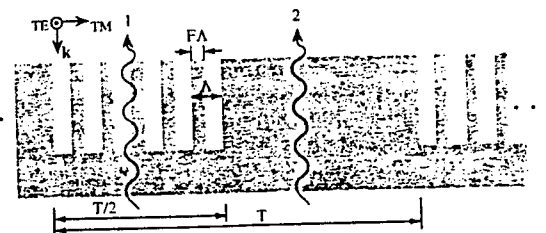


Fig. 1. FBCGH design: one period in a FBCGH.

than the period of the HSFG, second-order effective medium theory¹³ (EMT) can be used to calculate the effective indices with high accuracy. The effective indices for TE and TM polarizations calculated with a second-order EMT are given by

$$n_{2TE} = \left\{ n_{0TE}^2 + \frac{1}{3} \left[\frac{\Lambda \pi F}{\lambda} (1 - F)(n^2 - n_0^2) \right]^2 \right\}^{1/2},$$

$$n_{2TM} = \left\{ n_{0TM}^2 + \frac{1}{3} \left[\frac{\Lambda \pi F}{\lambda} n_{0TM} n_{0TE} \right. \right. \\ \left. \left. \times (1 - F) \left(\frac{1}{n^2} - \frac{1}{n_0^2} \right) \right]^2 \right\}^{1/2}, \quad (2)$$

where

$$n_{0TE} = [Fn^2 + (1 - F)n_0^2]^{1/2},$$

$$n_{0TM} = \left[\frac{(n_0 n)^2}{Fn_0^2 + (1 - F)n^2} \right]^{1/2}$$

are the effective indices calculated with zero-order EMT, Λ is the period of the HSFG, F is the grating fill factor of the HSFG (defined as the ratio between the width of the unetched portion within one period of grating to the grating period Λ ; see Fig. 1), and n and n_0 are the refractive indices of the two materials that form the HSFG. We chose GaAs as the substrate; therefore $n = n_s = 3.37$ (index of GaAs at 1.55 μm) and $n_0 = n_{\text{air}} = 1$. In general, $n_{TE} > n_{TM}$. To design a diffractive polarization beam splitter, we implement a binary phase grating for TE polarization (i.e., $\Phi_{TE} = \pi$) and without affecting the TM polarization (i.e., $\Phi_{TM} = 2\pi$).

Once the reconstruction wavelength λ is chosen, the period of the HSFG, Λ , can be determined. This Λ should be large enough to facilitate the fabrication and small enough not to cause higher than the 0th propagating diffraction orders. From our RCWA simulation we found that

$$\Lambda \leq \lambda/n_s \quad (3)$$

is a useful criterion.⁹ Thus we only need to find the grating fill factor F and the etch depth d to design the element. First, we determine F . From Eqs. (1) with our design parameters $\Phi_{TE} = \pi$ and $\Phi_{TM} = 2\pi$ we have

$$\frac{n_s - n_{TE}}{n_s - n_{TM}} = \frac{\Phi_{TE}}{\Phi_{TM}} = 1:2. \quad (4)$$

Substitute n_{TE} and n_{TM} from Eqs. (2) into Eq. (4); choose operating wavelength $\lambda = 1.55 \mu\text{m}$ and HSFG period $\Lambda = 0.3 \mu\text{m}$. By solving the resultant Eq. (4) we find the grating fill factor $F = 0.3509$. With this fill factor and other parameters, the corresponding effective refractive indices from Eqs. (2) are found to be

$$n_{2TE} = 2.309, \quad n_{2TM} = 1.2447.$$

Finally, we find from Eqs. (1) the required etch depth of the HSFG, $d = 0.728 \mu\text{m}$.

To ensure the accuracy of this design we also simulate the phase delay introduced by the HSFG, using a RCWA.^{9,10} In the RCWA a single period of a surface relief grating is divided into a large number of planar layers. The optical fields are formulated in terms of spatial harmonics by Fourier series expansions of the dielectric constant of each layer. Boundary conditions are matched and energy conservation law is employed to solve the resultant coupled diffraction equations. In our simulation we only try to calculate the phase delay caused by HSFG to confirm the results that we obtained by using the EMT. The actual diffraction efficiency of a FBCGH is estimated later by scalar diffraction theory. The grating parameters are the same as those given above. The simulation indicates that the phase delay introduced by a 0.73- μm -thick HSFG is 2.154π for TE polarization and 1.190π for TM. A GaAs layer of the same thickness without HSFG introduces 3.178π phase delay. Thus the designed grating will have a 1.024π phase difference between the HSFG pixel and an unetched pixel for TE polarization and 1.987π for TM polarization. This simulation shows the validity of our design. It also indicates that the EMT, if used carefully, can be used in designing FBCGH elements.

Following this design, we fabricated a diffractive structure upon a (100)-cut GaAs substrate, using electron beam lithography and dry etching techniques.⁸ The total area of the element was $100 \mu\text{m} \times 100 \mu\text{m}$. The period of the binary phase diffractive grating T was $10 \mu\text{m}$. The period of HSFG Λ was $0.3 \mu\text{m}$, and the fill factor of the HSFG F was 0.35. The fabricated element has an etch depth of $0.75 \mu\text{m}$ for the HSFG. Figure 2 shows a scanning electron micrograph of the fabricated element.

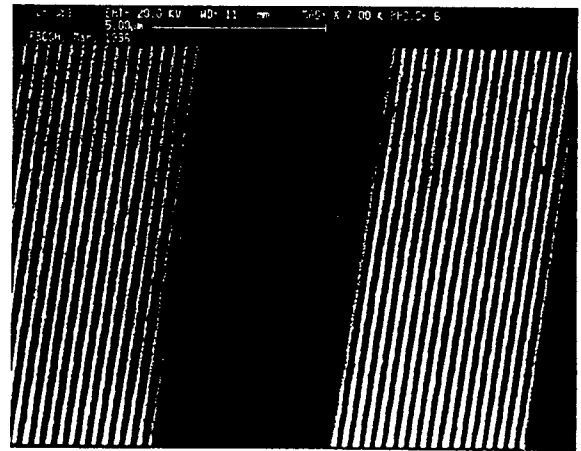


Fig. 2. Scanning electron micrograph of the fabricated FBCGH.

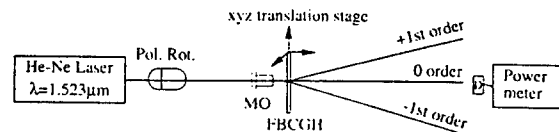


Fig. 3. Schematic of the experimental evaluation of the fabricated FBCGH.

Table 1. Measured Results of the Fabricated FBCGH^a

Performance	0th Order	1st Order	-1st Order
TE efficiency	0.86% (0.0%)	41.4% (40.5%)	44.2% (40.5%)
TM efficiency	75.5% (100%)	0.15% (0%)	0.44% (0%)
Polarization contrast ratio	88.2:1	275:1	99.2:1

^aThe results calculated by scalar diffraction theory for a binary phase level diffractive optical element are given in parentheses for comparison.

We evaluated the fabricated element with a He-Ne laser (Melles Griot) operating at 1.523 μm , using the setup shown schematically in Fig. 3. The beam was focused onto the FBCGH by a low-power (6 \times) microscope objective (MO). A Ge detector was used to measure the far-field diffraction patterns. The polarization state of the beam incident upon the FBCGH was controlled with a polarization rotator (Pol. Rot.).

In the binary phase FBCGH reconstruction stage we anticipate observing +1 and -1 propagating diffraction orders. In our characterization experiments we observed only two spots on the IR phosphor viewing card under TE polarization and one spot under TM polarization, although higher orders do exist and can be detected with a photodetector. We can optimize the distance between the microscope objective and the FBCGH by minimizing the measured energy diffracted into the 0th diffraction order at TE-polarized illumination. The measured diffraction efficiency, excluding reflection, and the polarization contrast ratios are summarized in Table 1. The diffraction efficiency of Table 1 was calculated as the ratio between the intensity measured at a certain diffraction order and that of the total light transmitted through the GaAs substrate without a FBCGH. These measured results show that the FBCGH has good polarization selectivity (large polarization contrast ratios) and diffraction efficiencies close to the theoretical limit. Note that the form-birefringent structure also serves as an antireflection coating, explaining the slightly higher measured diffraction efficiencies compared with that predicted by scalar diffraction theory for a binary phase element (40.5%). The expected results calculated with scalar diffraction theory are also listed in the table for comparison. The slight asymmetry between the efficiencies of ± 1 st diffraction orders is due to imperfect normal incidence.

In conclusion, we have designed, fabricated, and evaluated a polarization-selective computer-generated hologram that uses form-birefringent nanostructures upon GaAs substrates. The element was designed by use of effective-medium theory and verified to

be valid by the rigorous vector field theory. The design and the experimental evaluations were found to be in good agreement. The fabricated element shows a large polarization contrast ratio (as large as 275:1) and high diffraction efficiencies (>40% for the first diffraction orders). Such an element may be useful in fabrication of compact and efficient free-space transparent photonic switching fabrics as well as packaging optoelectronic devices and systems.

The research conducted at the University of California, San Diego, is funded by the National Science Foundation U.S. Air Force Office of Scientific Research, and Rome Laboratories and that at Caltech is funded by the National Science Foundation.

References

1. F. Xu and Y. Fainman, *Appl. Opt.* **34**, 256 (1995).
2. J. Ford, F. Xu, K. Urquhart, and Y. Fainman, *Opt. Lett.* **18**, 456 (1993).
3. N. Nieuborg, C. Van de Poel, A. Kirk, H. Thienpont, and I. Veretennicoff, in *Optical Computing*, Vol. 10 of 1995 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1995) p. 124.
4. F. Xu, R. Tyan, J. Ford, and Y. Fainman, *Opt. Lett.* **21**, 516 (1996).
5. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, London, 1986).
6. D. C. Flanders, *Appl. Phys. Lett.* **42**, 492 (1983).
7. P. Kipfer, M. Collischon, H. Haidner, J. Schwider, N. Streibl, and J. Lindolf, *Opt. Eng.* **33**, 79 (1994).
8. F. Xu, R. C. Tyan, P. C. Sun, C. C. Cheng, A. Scherer, and Y. Fainman, *Opt. Lett.* **20**, 2457 (1995).
9. I. Richter, P. C. Sun, F. Xu, and Y. Fainman, *Appl. Opt.* **34**, 2421 (1995).
10. E. N. Glytsis and T. K. Gaylord, *Appl. Opt.* **31**, 4459 (1992).
11. F. T. Chen and H. G. Craighead, *Opt. Lett.* **20**, 121 (1995).
12. S. Aoyama and T. Yamashita, *Proc. SPIE* **1545**, 241 (1991).
13. S. M. Rytov, *Sov. Phys. JETP* **2**, 466 (1956).

**A. 7. Design, fabrication, and characterization of form-birefringent
multilayer polarizing beam splitter**

R.C. Tyan, A.A. Salvekar, H.P Chuo, C.C. Cheng, A. Scherer, P.C. Sun, F. Xu and Y.
Fainman

JOSA A, Vol. 14, No. 7, July, 1997.

Design, fabrication, and characterization of form-birefringent multilayer polarizing beam splitter

Rong-Chung Tyan

Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, California 92093

Atul A. Salvekar, Hou-Pu Chou, Chuan-Cheng Cheng, and Axel Scherer

Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125

Pang-Chen Sun, Fang Xu, and Yeshayahu Fainman

Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, California 92093

Received September 13, 1996; accepted November 21, 1996

Polarizing beam splitters that use the anisotropic spectral reflectivity (ASR) characteristic of high-spatial-frequency multilayer binary gratings have been designed, fabricated, and characterized. Using the ASR effect with rigorous coupled-wave analysis, we design an optical element that is transparent for TM polarization and reflective for TE polarization at an arbitrary incidence angle and operational wavelength. The experiments with the fabricated element demonstrate a high efficiency ($>97\%$), with polarization extinction ratios higher than 220:1 at a wavelength of $1.523\ \mu\text{m}$ over a 20° angular bandwidth by means of the ASR characteristics of the device. These ASR devices combine many useful characteristics, such as compactness, low insertion loss, high efficiency, and broad angular and spectral bandwidth operations. © 1997 Optical Society of America [S0740-3232(97)01107-1]

1. INTRODUCTION

Numerous optical information processing and imaging systems employ different polarization states to increase the information bandwidth and to reduce the cross talk between different channels. Some application examples include free-space optical switching networks,¹⁻³ read-write magneto-optical data storage systems,^{4,5} and polarization-based imaging systems.^{6,7} In these systems, a polarizing beam splitter (PBS) is an essential element for separating two orthogonally polarized light beams. Most of the applications require that the PBS providing high extinction ratios tolerate a wide angular bandwidth for high-resolution imaging, a broad wavelength range for operation with broadband sources, and a compact size for effective packaging. Conventional PBS based on either natural crystal birefringence (e.g., Wollaston prisms) or polarization selectivity of multilayer structures (e.g., PBS cubes) do not meet these requirements. The Wollaston prism requires a large thickness to generate enough walk-off distance between the two orthogonal polarizations owing to the intrinsically small birefringence of naturally anisotropic materials. An alternative design of a Wollaston-type prism takes advantage of form-birefringent materials that possess birefringence several times larger than that of natural birefringent materials, reducing the thickness considerably.⁸ However, the fabrication of such a structure is a tedious and complicated

process, since thousands of layers of thin-film slab need to be fabricated. PBS cubes are easier to fabricate, but they provide good extinction ratios only in a narrow angular bandwidth for a limited spectral range.⁹ Other designs that utilize form-birefringent high-spatial-frequency surface-relief gratings¹⁰⁻¹² and a single-layer-coated dielectric slab¹³ have been proposed to reduce the size of the components, to solve the material compatibility problem, and to simplify the fabrication process. However, they usually suffer from low efficiency, low extinction ratio, small angular bandwidth, and operation in a limited wavelength range.

Previously¹⁴ we introduced a new PBS device that uses the unique properties of anisotropic spectral reflectivity (ASR) characteristics of a high-spatial-frequency multilayer binary grating. The ASR mechanism is based on combining the effects of the form birefringence of a high-spatial-frequency grating (i.e., grating period is much smaller than the wavelength of the incident field) with the resonant reflectivity of a multilayer structure. With our approach, the angular field and the wavelength range have been largely increased compared with conventional PBS devices. Many packaging and material compatibility problems have also been resolved with this new design. Some interesting characteristics of the element with ASR characteristics cannot be found in a conventional PBS component. For instance, when our ASR de-

vice is designed to operate with normally incident light, it acts as a highly efficient polarization-selective mirror.

In this paper we report detailed design, fabrication, and experimental characterization of such a PBS based on ASR properties. In Section 2 we describe the principle of the ASR effect of a high-spatial-frequency multilayer binary grating. We employ the effective-medium theory (EMT)¹⁵ to explain intuitively why high-spatial-frequency multilayer binary gratings possess such characteristics. In Section 3 we introduce the design methodology of the PBS that employs such ASR properties. Here rigorous coupled-wave analysis (RCWA)¹⁶ tools are used to optimize¹⁷ the design of the PBS. We also investigate the angular and the wavelength dependence of the ASR PBS. In Section 4 the effects of various fabrication errors on performances of the PBS are studied. In Section 5 we discuss the fabrication techniques employed to make our PBS, and we present experimental characterization results. We evaluate our PBS design in terms of polarization extinction ratio and efficiency for operation with waves of wide angular bandwidth. We also compare the experimental results with the numerical predictions. The summary and the directions for future research are provided in Section 6.

2. PRINCIPLES OF ANISOTROPIC SPECTRAL REFLECTIVITY

For reviewing properties of wave propagation in stratified media, consider a multilayer structure formed on a sub-

strate by deposition of alternating layers of isotropic dielectric materials with high and low indices of refraction n_h and n_l , respectively. Such a structure exhibits high reflectivity in a wide spectral bandwidth. When the thickness of each layer corresponds to a quarter-wave optical thickness for a selected wavelength, a high-reflectivity spectral band will roughly center at that wavelength. For example, Fig. 1(a) shows a 15-layer quarter-wave structure made of Si and SiO₂ with refractive indices of 3.48 and 1.44,¹⁸ respectively, for an incident wavelength of 1.523 μm . A high-reflectivity spectral band is clearly shown in Fig. 1(a) near that wavelength. One can increase the reflectivity of such a quarter-wave layered structure by increasing the value of the ratio n_h/n_l and the total number of layers.¹⁹ Furthermore, larger values of the ratio n_h/n_l also increase the spectral bandwidth of high reflectivity.

For an optical field at a normal angle of incidence, a multilayer structure that is made of isotropic dielectric materials presents identical reflectivity spectra for any two orthogonal linear polarizations. This occurs because of the symmetry of the structure for a normally incident wave. Therefore the structure cannot be used to separate normally incident fields by polarization. By substituting one of the isotropic materials that are used to form the multilayer structure with birefringent materials, we can create a new multilayer structure that will possess reflectivity spectral bands centered at different wavelengths for the two orthogonal polarizations at normal incidence. It is illustrated in Fig. 1(b) that once one of the

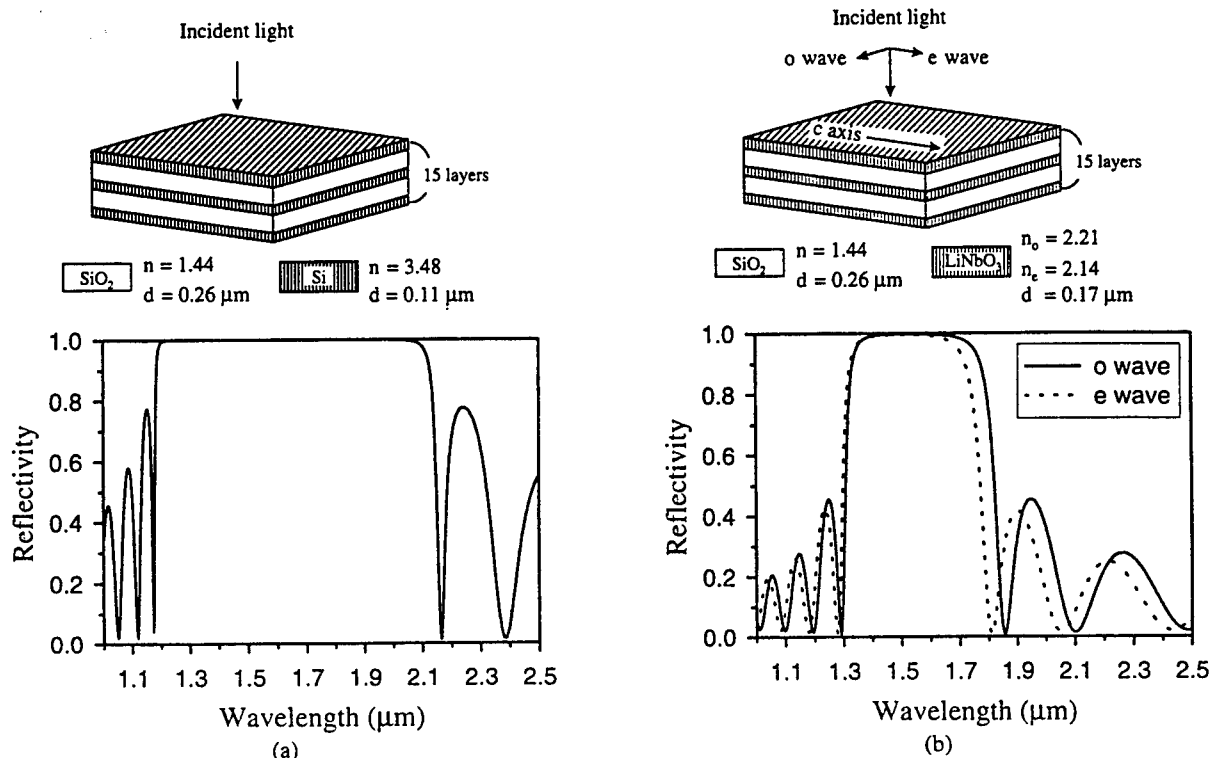


Fig. 1. (a) Schematic diagram of a 15-layer quarter-wave structure constructed of two isotropic materials (Si and SiO₂); a plot of its spectral reflectivity is also shown. The thickness of Si and SiO₂ layers are set as the quarter-wave optical thickness of the wavelength of 1.523 μm . (b) Same as (a), with Si being replaced by anisotropic material, LiNbO₃. The thickness of the LiNbO₃ layer is set as the quarter-wave optical thickness corresponding to the refractive index for the ordinary wave.

isotropic materials (say, Si) is replaced by an anisotropic material the spectral reflectivity bands are separated [in this example we disregard the material incompatibility problems and choose LiNbO₃ with refractive indices of 2.21 and 2.14 (Ref. 20) for ordinary and extraordinary waves, respectively]. We call these phenomena ASR. Unfortunately, multilayer structures consisting of natural anisotropic materials cannot easily be fabricated. Furthermore, since natural anisotropic materials possess small birefringence, this separation will be small. With our approach based on a high-spatial-frequency multilayer binary grating, the separation between reflectivity spectral bands can increase considerably owing to the exceptionally large anisotropy²¹ that can be obtained with a form-birefringent nanostructure.

Form-birefringence effects²² appear in high-spatial-frequency gratings constructed of isotropic dielectric materials. Because of the geometric asymmetry of the grating structure, the two orthogonally polarized optical fields, one parallel to the grating grooves (called the TE field) and the other perpendicular to the grating grooves (called the TM field), encounter different boundary conditions, resulting in distinct effective indices of refraction. On propagation through the grating structure, the TE and the TM fields will acquire a relative phase difference similar to that obtained in natural anisotropic materials. This similarity takes place because the subwavelength structure can be designed such that only the zeroth diffraction order will propagate, while all the higher diffraction orders will become evanescent. Such a high-spatial-frequency grating at the boundary between two isotropic materials can be seen as an equivalent thin film of anisotropic material.

For normally incident TE- and TM-polarized optical waves, the effective indices of refraction of a surface-relief high-spatial-frequency binary grating can be estimated from the second-order EMT¹⁵:

$$n_{TE}^{(2)} = \left[n_{TE}^{(0)2} + \frac{1}{3} \left(\frac{\Lambda}{\lambda} \right)^2 \pi^2 F^2 (1-F)^2 (n_{III}^2 - n_I^2)^2 \right]^{1/2}, \quad (1)$$

$$n_{TM}^{(2)} = \left[n_{TM}^{(0)2} + \frac{1}{3} \left(\frac{\Lambda}{\lambda} \right)^2 \pi^2 F^2 (1-F)^2 \left(\frac{1}{n_{III}^2} - \frac{1}{n_I^2} \right)^2 \right. \\ \left. \times n_{TE}^{(0)2} n_{TM}^{(0)6} \right]^{1/2}, \quad (2)$$

where F is the duty cycle of the grating defined by $F = w/\Lambda$, with w being the width of the binary grating [see Fig. 2(a)]; Λ is the grating period; λ is the wavelength of the incident wave; n_I and n_{III} are the indices of air and the grating material, respectively; and $n_{TE}^{(0)} = [Fn_{III}^2 + (1-F)n_I^2]^{1/2}$ and $n_{TM}^{(0)} = \{n_{III}^2 n_I^2 / [Fn_I^2 + (1-F)n_{III}^2]\}^{1/2}$ are the effective indices of refraction for TE and TM waves, respectively, provided by the zero-order EMT.²² When a high-spatial-frequency grating is formed in a multilayer structure made of two isotropic materials, the composition becomes an artificial anisotropic multilayer structure that will possess ASR characteristics. One can fabricate such an element by etching a high-spatial-frequency binary grating directly into a multilayer mirror structure.

Figure 2(a) shows an example of a high-spatial-frequency multilayer binary grating. The two isotropic materials used for constructing the multilayer structures are SiO₂ and Si, with refractive indices of 1.44 and 3.48, respectively, for an operating wavelength of 1.523 μ m. The SiO₂ and Si materials are chosen because of their fabrication compatibility and because of low absorption coefficients in the near-infrared region, which ensure a low insertion loss of the device. For operation as a form-birefringent zeroth-diffraction-order grating, we set the grating period to be equal to 0.6 μ m, with the duty cycle of $F = 0.5$. Using second-order EMT [Eqs. (1) and (2)], we

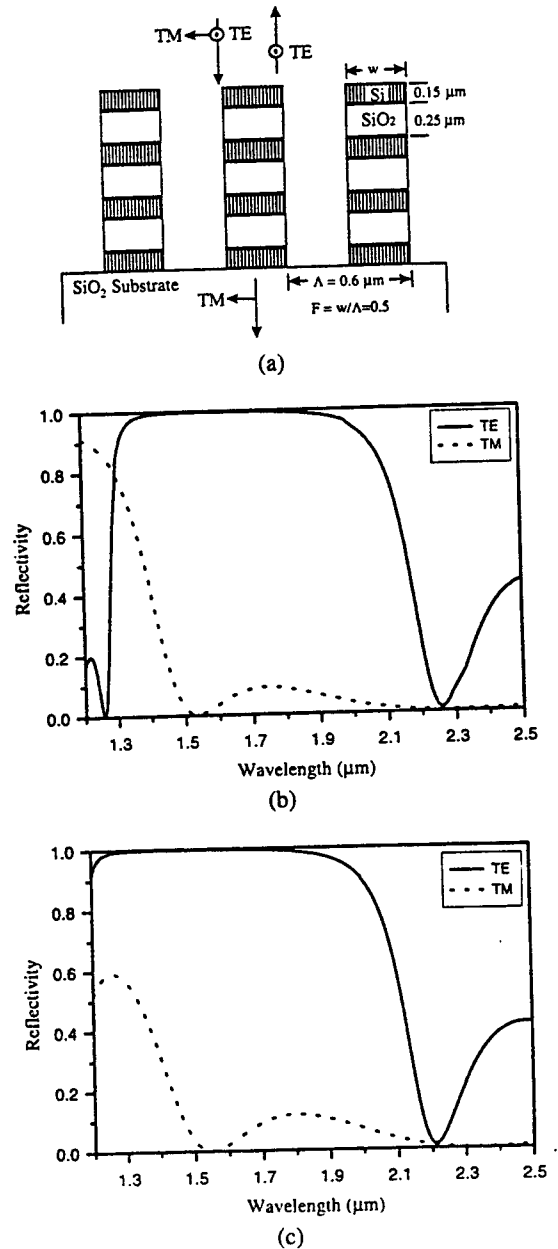


Fig. 2. (a) Schematic diagram of a seven-layer ASR PBS designed for light at normal incidence. The center operating wavelength is 1.523 μ m. The design parameters are indicated in the figure. (b) EMT and (c) RCWA results of the reflectivity for TE- and TM-polarized waves versus wavelength.

obtain the following effective refractive indices for the two materials: $n_{\text{TE}, \text{Si}}^{(2)} = 3.24$, $n_{\text{TE}, \text{SiO}_2}^{(2)} = 1.25$ and $n_{\text{TM}, \text{Si}}^{(2)} = 1.72$, $n_{\text{TM}, \text{SiO}_2}^{(2)} = 1.18$. The effective indices of both materials used in the multilayer structure are larger for TE polarization than for TM polarization. This indicates that in the spectral domain the reflectivity band for TE polarization will be centered at a longer wavelength as compared with that for TM polarization, i.e., we will clearly observe a large ASR effect. The separation between the two reflectivity spectral bands has been dramatically increased over that possible with natural anisotropic materials.

The ASR characteristic of a multilayer form-birefringent binary grating is the essential property needed to realize the ASR PBS. By using the theory of optical wave propagation in a stratified medium²³ that is characterized by second-order EMT, we show that the ASR property can be used to create a compact PBS. Figure 2(b) shows the spectral reflectivities of the seven-layer grating illustrated in Fig. 2(a). It can clearly be seen that near a wavelength of $1.523 \mu\text{m}$ the grating is transparent for TM polarization and reflective for TE polarization. Therefore the two orthogonal polarization components of the light wave can be separated to propagate in opposite directions with efficiencies close to 100%.

3. DESIGN AND MODELING OF THE POLARIZING BEAM SPLITTER

In Section 2 we used second-order EMT for an intuitive explanation of the ASR characteristic of high-spatial-frequency gratings. However, previous studies^{17,24,25} have shown that, to accurately model devices in which the grating period is compatible with the wavelength, a rigorous method must be employed. In our case the grating period is approximately one third of a wavelength. Our modeling method is to use EMT results as an initial estimate of device parameters and then use a rigorous method to modify the design for optimal performance. The modeling tool that we used is RCWA, which we have experimentally verified for its accuracy in modeling of surface-relief-type grating.^{21,26}

The design procedure is described by the following two steps: First, we use second-order EMT to calculate the effective indices of TE- and TM-polarized light in each layer of the grating at the desired center operating wavelength. The grating period is set to be small enough that only the zeroth diffraction order can propagate. From our previous study¹⁷ we know that a high-spatial-frequency grating can possess the largest effective-index difference for waves with orthogonal polarizations when the duty cycle is near 0.5. Therefore we set the duty cycle $F = 0.5$ as an initial value. An interesting characteristic is that the value of the effective-index ratios for TE-polarized light is larger than that for TM-polarized light. For example, effective-index ratios of the grating illustrated in Fig. 2(a), $(n_h/n_l)_{\text{TE}}$ and $(n_h/n_l)_{\text{TM}}$, are 2.59 and 1.46, respectively. This indicates that, to achieve the same reflectivity, the number of layers required for TE polarization will be less than that required for TM polarization. To minimize the number of layers needed to

achieve a desired performance, we choose to maximize reflectivity for TE-polarized light.

In the next step we allow each layer to have a quarter-wave optical thickness based on the effective index for TE-polarized light. These parameters are now used as the basis for an optimum design by RCWA. The thickness and the duty cycle of a high-spatial-frequency grating have been shown¹⁷ to be the most important parameters that will affect the phase difference between the two orthogonally polarized waves. Optimization is performed by incremental variation of the thickness of the layers of the grating to obtain the highest extinction ratio at the operational wavelength. To achieve broad reflectivity spectral bands, we use high-refractive-index materials for both the first and the last layers in the structure.¹⁹

Figure 2(c) shows the RCWA results of TE and TM reflectivities as a function of the wavelength for the seven-layer ASR PBS shown in Fig. 2(a). The PBS is designed for normally incident optical fields at a center operating wavelength of $1.523 \mu\text{m}$. As expected, the reflectivity of TE-polarized light is higher than that of TM-polarized light. The TE reflectivity spectral band is broader and is centered at a longer wavelength than is that of the TM wave. This ASR property cannot be accomplished with an isotropic multilayer structure for a normally incident optical field. Since this PBS is made of nonabsorbing materials (at those operating wavelengths), there is basically no insertion loss, and the efficiency is nearly 100%. Looking at Fig. 2(c), we can see that both the TE reflection efficiency and the TM transmission efficiency are higher than 99% at the operating wavelength of $1.523 \mu\text{m}$. Also, the transmission polarization extinction ratios (defined as the ratio of transmittance of TM-polarized light to that of TE-polarized light) remain high over a broad spectral range ($>200:1$ over a 200-nm range). However, the reflection polarization extinction ratios (defined as the ratio of reflectance of TE-polarized light to that of TM-polarized light) are extremely high over a small spectral range ($>1000:1$ over a 20-nm range). These unique features can be employed in constructing either broadband low-insertion-loss normal-incidence polarizers or highly efficient polarization-selective mirrors for microlaser cavities. Other ASR PBS designs following the design procedure mentioned above for a center operating wavelength of $1.3 \mu\text{m}$ can be found in our previous paper.¹⁴

To separate the path of the reflected wave from that of the incident wave, we also investigate an off-axis geometry [Fig. 3(a)]. Figure 3(b) shows the numerical results of the reflectivity versus the wavelength of the slanted incidence optical wave from a five-layer grating. For an incident wavelength of $1.523 \mu\text{m}$, both the TE reflection efficiency and the TM transmission efficiency are higher than 99%, and the polarization extinction ratios for reflection and transmission are better than 800:1 and 300:1, respectively. This slanted incidence arrangement offers two advantages: (1) Reflectivity from each layer for TE polarization is increased; thus only five layers were needed to achieve performance similar to those of the seven-layer design for normally incident light; and (2) the sidelobe for the TM reflectivity is flattened, allowing op-

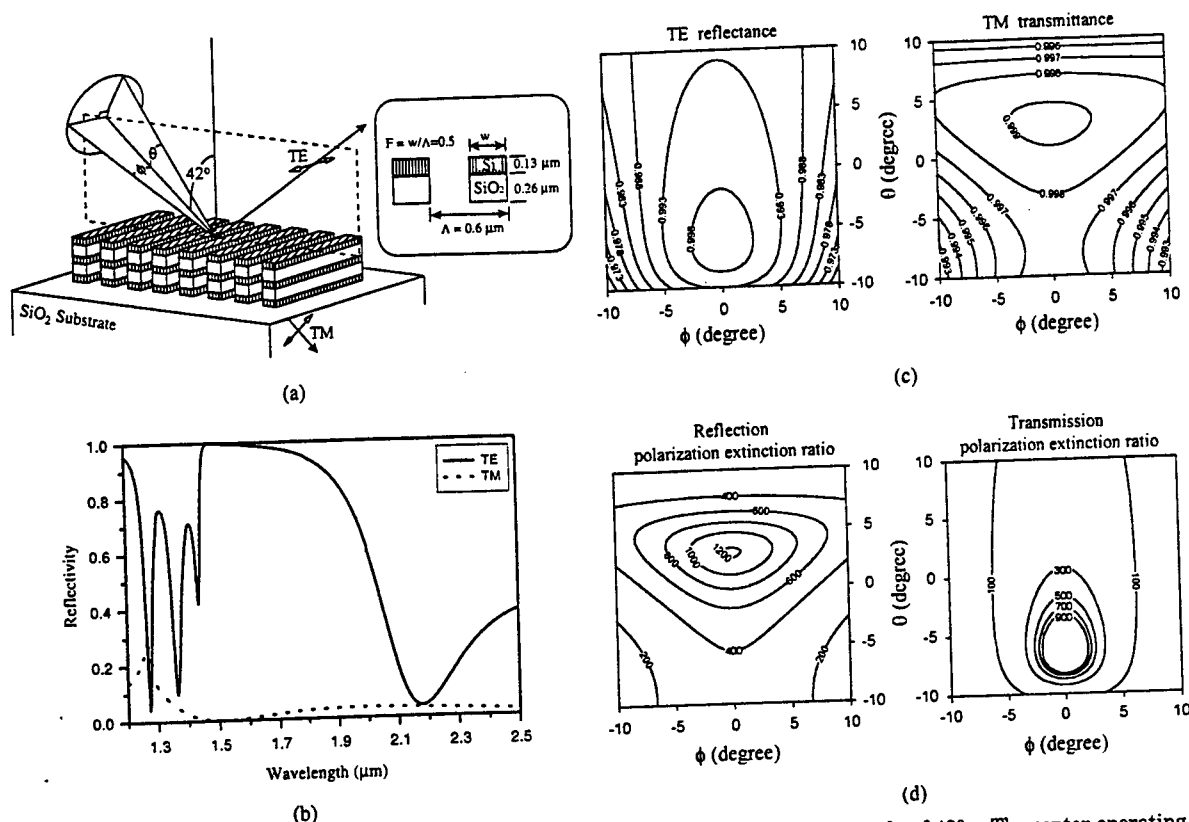


Fig. 3. (a) Schematic diagram of a five-layer ASR PBS operating with waves incident at an angle of 42° . The center operating wavelength is $1.523 \mu\text{m}$. The design parameters are indicated in the figure. (b) RCWA results for the reflectivity of TE- and TM-polarized waves versus wavelength. The angles of θ and ϕ are zeros in this calculation. Contour plots of (c) TE reflectance and TM transmittance and (d) the reflection and transmission polarization extinction ratio versus angles (ϕ, θ) are as defined in Fig. 3(a). The angles of incident waves are varied to span an angular bandwidth of $\pm 10^\circ$ in both θ and ϕ directions near the initial bias angle of 42° .

eration of the beam splitter in an even broader spectral range. One can further improve the efficiencies and the extinction ratios of the PBS by fine tuning the grating period and the duty cycle or by adding more layers to the structure.

Various applications require that the optical components in the system have abilities to maintain good performance within a large two-dimensional (2D) angular bandwidth and broad spectral range of light waves. Conventional PBS's provide good extinction ratios, usually in a narrow angular bandwidth for a limited wavelength range. Most of the polarization-sensitive diffractive optical elements also lack these abilities. Our PBS, however, is shown to provide good performances for optical signals that have a wide 2D angular bandwidth, as well as a broad spectral range. As illustrated in Fig. 3(a), the angles of incidence are varied to span an angular bandwidth of $\pm 10^\circ$ in both θ and ϕ directions defined near the initial 42° bias angle. The results shown in Fig. 3(c) indicate that, at a wavelength of $1.523 \mu\text{m}$, the TE reflection efficiencies are higher than 99.2% and 98% in the 5° and 10° angular bandwidth cones, respectively, and the TM transmission efficiencies are higher than 99.6% in both these angular ranges. The polarization extinction ratios [see Fig. 3(d)] are better than 400:1 and 200:1 for reflection in the 5° and the 10° angular bandwidth cones, respectively. For the transmission, the polarization ex-

ting ratios are smaller but still better than 130:1 and 50:1 in the 5° and the 10° angular bandwidth cones, respectively. These results show that a wide 2D angular bandwidth, as well as a broad spectral range, of operation is possible with this design. Furthermore, the property of high and uniform efficiency makes the ASR PBS suitable for many imaging systems applications.

4. STUDIES OF FABRICATION TOLERANCES OF THE POLARIZING BEAM SPLITTER

In this section we investigate the effect of fabrication errors (e.g., etch depth error, grating profile error, and duty cycle error) on the performance of the ASR PBS. We first investigate the effect of etch depth fabrication error, distinguishing between the two types of errors: an under-etched grating [see Fig. 4(a), i.e., part of the last thin Si layer remains on the SiO_2 substrate], and an overetched grating [see Fig. 4(b), i.e., the substrate is slightly etched]. The results of our simulation show that the reflectivity for TM polarization is more sensitive to the underetching error than is that for TE polarization [see Fig. 4(a)]. The remaining thin layer of Si will increase the TM reflectivity, since Si is a dense medium, resulting in a large reduction in the TM transmission efficiency. This also decreases the reflection polarization extinction ratio.

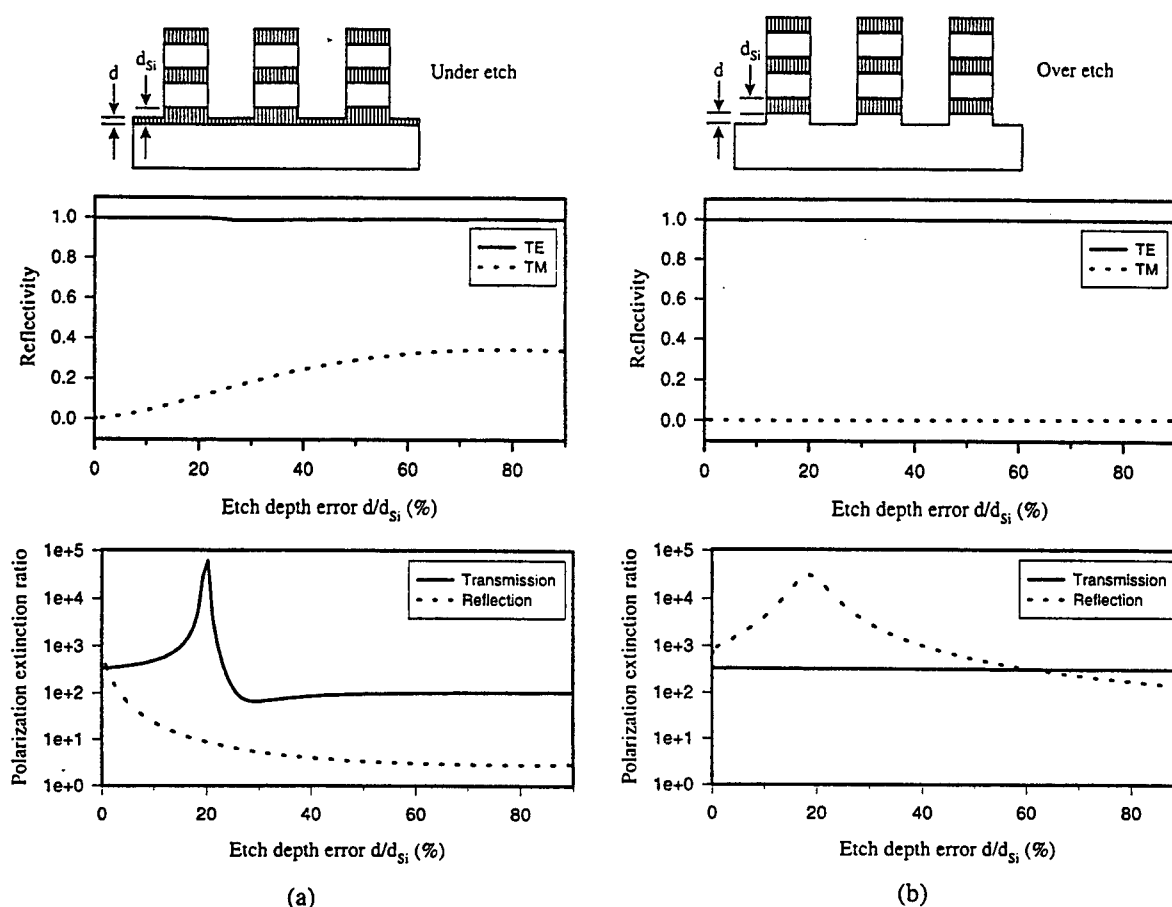


Fig. 4. Effects of (a) underetching and (b) overetching fabrication error on reflectivity and extinction ratios for TE- and TM-polarized light.

In contrast, Fig. 4(b) shows that overetching error has almost no effect on the efficiency of both TE- and TM-polarized waves. We conclude that an accurate etch depth control to avoid underetching is necessary to attain good performance from these devices.

Another possible fabrication error is the duty ratio error of the high-spatial-frequency grating. We use the same design parameters as for the PBS shown in Fig. 3(a), except that the duty cycle F is varied continuously from 0.3 to 0.6 [see Fig. 5(a)]. The numerical results indicate that the reflectances for both polarizations stay approximately the same for the different duty cycles ranging from 0.3 to 0.55 [see Fig. 5(b)]. The abrupt decrease of the TE reflectance at the duty cycle of 0.57 may be due to the energy coupling into the guided waves propagating along the multilayer structure. We can observe from Fig. 5(c) that, when the duty cycle of the grating is approaching 0.54, the transmission polarization extinction ratios are briefly increasing. Since the polarization extinction ratio is a ratio of a large value to an extremely small value, a minor variation of the small value will cause an abrupt change of the ratio. These resonance phenomena can also be seen in the other figures (see, e.g., Fig. 4).

The last fabrication error that we investigated is the shape of the grating profile. Again, we used the same basic design shape as that of the PBS shown in Fig. 3(a).

We vary the shape of the grating profile by holding the width W of the bottom of the grating constant as the top is varied as $W - \Delta W$ to form a symmetric trapezoidal profile, as illustrated in Fig. 6(a). The results shown in Fig. 6(b) indicate that the efficiencies of the PBS will remain nearly 100% over a fairly large grating profile error. The polarization extinction ratios [see Fig. 6(c)] slowly decrease as the shape changes from rectangular to trapezoidal.

In general, the ASR PBS are sensitive to underetching fabrication errors. However, these elements are relatively immune to the effects of fabrication errors in duty cycle, grating profile, and overetching. This indicates a rather large fabrication error tolerance of the ASR PBS. In Section 5 we discuss fabrication and experimental verification of such ASR PBS.

5. FABRICATION AND EXPERIMENTAL CHARACTERIZATION OF THE POLARIZING BEAM SPLITTER

Fabrication of the ASR PBS in the visible spectral range is a challenging task because of the requirement to fabricate a grating with a subwavelength grating period. However, for near-infrared range of operation, fabrication of the structure is practically possible. For example, in

our design the total grating depth of five layers is $0.91 \mu\text{m}$, with a grating period of $0.6 \mu\text{m}$ and a duty cycle of 0.5, resulting in a grating aspect ratio of approximately 3:1, which is within the fabrication capabilities of modern microfabrication technology.

The fabrication procedures of the ASR PBS are shown schematically in Fig. 7. First, five layers of Si and SiO_2 , with thickness of 0.13 and $0.26 \mu\text{m}$, respectively, were sputtered alternatively onto a SiO_2 substrate to form a one-dimensional (1D) dielectric mirror base. The thickness of each layer was controlled with an accuracy of better than 5 nm. 30-kV high-voltage *e*-beam lithography was then used to define a high-resolution grating with a period of $0.6 \mu\text{m}$ and a duty cycle of 0.5 over a square area of $50 \mu\text{m} \times 50 \mu\text{m}$ on a $0.4\text{-}\mu\text{m}$ -thick PMMA layer that was thicker than the most conventional *e*-beam lithography resists [see Fig. 7(a)]. The PMMA pattern was developed for 60 s in a 3:7 mixture of cellulose and methanol, resulting in a structure shown in Fig. 7(b). Afterward, $0.1 \mu\text{m}$ of chrome was deposited [see Fig. 7(c)],

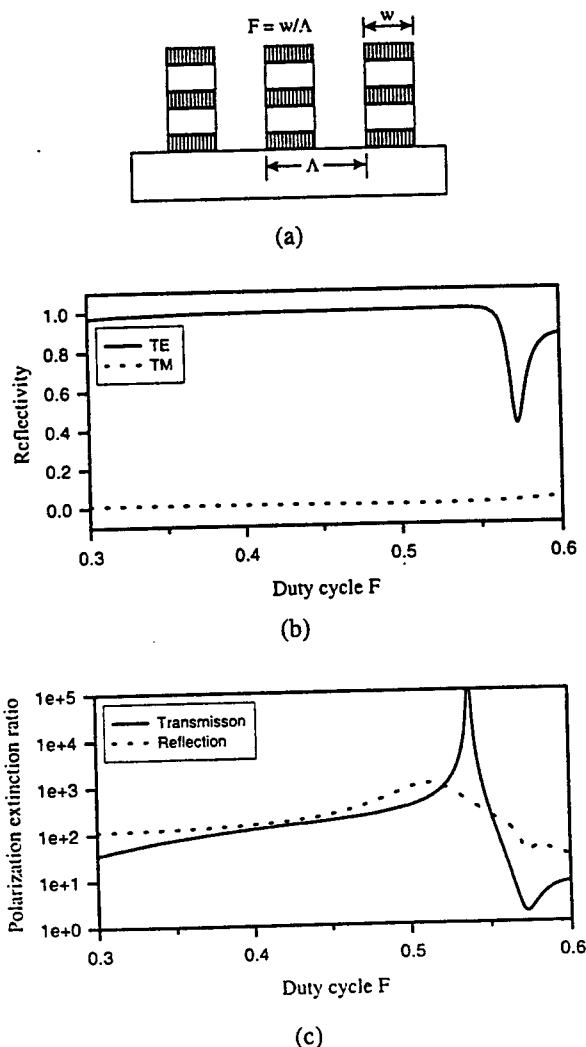


Fig. 5. (a) Diagram defining duty cycle error. Effect of duty cycle error on (b) reflectivity and (c) extinction ratios of TE- and TM-polarized light.

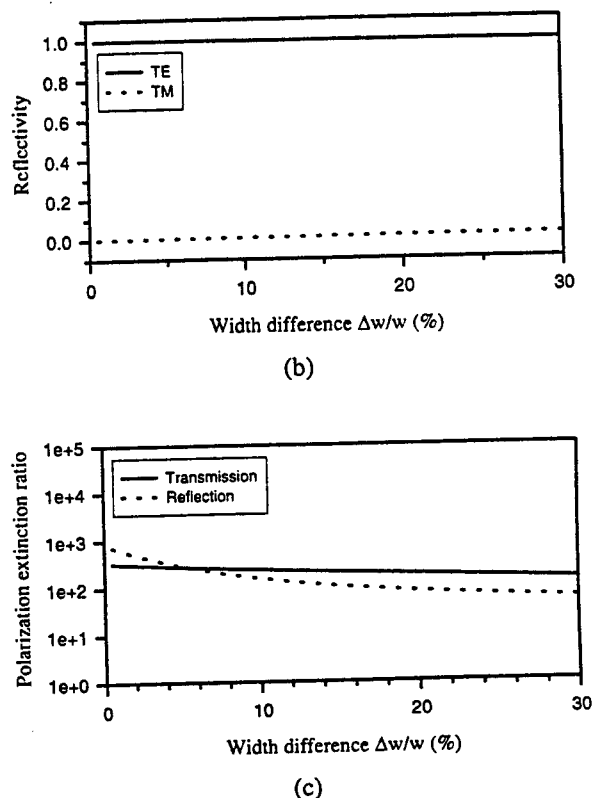
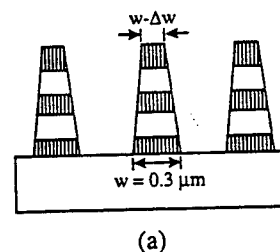


Fig. 6. (a) Diagram defining grating shape error causing trapezoidal profile. Effect of grating profile error on (b) reflectivity and (c) extinction ratios for TE- and TM-polarized light.

and the PMMA pattern was lifted off in acetone to generate the mask [see Fig. 7(d)]. With the durable chrome mask, the periodic pattern was transferred through the SiO_2 and the Si layers by reactive ion etching with C_2F_6 and $\text{NF}_3/\text{CCl}_2\text{F}_2$, respectively [see Fig. 7(e)]. The etch rate of SiO_2 and Si was 25 and 100 nm/min, respectively. Finally, the remained chrome mask was removed by wet etching [see Fig. 7(f)]. The fabricated PBS was inspected under a scanning electron microscope (SEM). The SEM's side view of the structure is shown in Fig. 8.

The fabricated PBS element was evaluated experimentally by means of a measurement setup shown schematically in Fig. 9. We used a polarized He-Ne laser source (Melles Griot) operating at a near-infrared wavelength of $1.523 \mu\text{m}$ with a 0.8-mW maximum output power. The laser beam is focused onto the $50 \mu\text{m} \times 50 \mu\text{m}$ aperture of the fabricated element with a low-power ($5\times$) microscope objective. The input polarization is controlled by a polarization rotator. Two Ge photodetectors (Newport

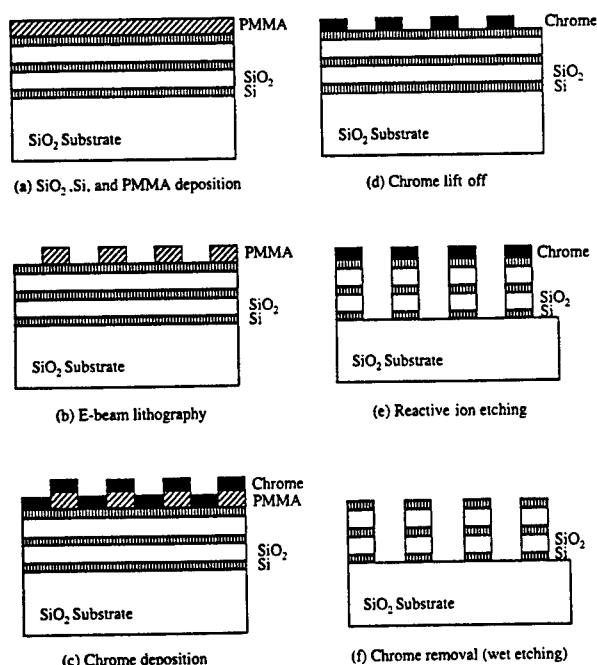


Fig. 7. Schematic diagram describing the fabrication procedures of ASR PBS on SiO_2 substrates. PMMA, poly(methyl methacrylate).

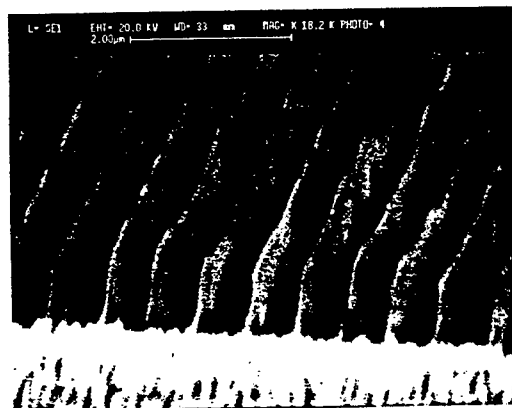


Fig. 8. SEM photograph of the fabricated ASR PBS. The PBS is fabricated on a SiO_2 substrate consisting of a multilayer structure of Si and SiO_2 with a thickness of 0.13 and 0.26 μm , respectively. The grating has a period of 0.6 μm , with a duty cycle of 0.5.

Model 818-IR) were used to measure the transmittance and the reflectance simultaneously. For the alignment purpose we also used another He-Ne laser operating at a visible wavelength of 0.6328 μm . The axis of the visible He-Ne laser was aligned to coincide with the optical axis of the infrared laser.

The measured efficiencies (i.e., the TM transmittance and the TE reflectance) and the measured polarization extinction ratios versus the incidence angles varying from 32° to 52° are shown in Figs. 10(b) and 10(c), respectively. The incidence angle is determined by the angle between the incident beam and the normal to the grating surface. The incident beam is lying in the plane perpendicular to

the grating grooves and parallel to the grating vector. In such an arrangement, the measured 1D efficiency and polarization extinction ratio curves correspond to those determined numerically from the vertical cross sections of the 2D contour plots shown in Fig. 3. The reflectance is measured in a range of 36–52° because of the practical constraints in the components used in our experimental setup. The experimental results show that the fabricated ASR PBS retains high-polarization extinction ratios over a large angular bandwidth (± 10 degree) from the designed incidence angle of 42°. The measured transmission polarization extinction ratios are higher than 220:1, with a maximum value of 830:1. For the reflection, the polarization extinction ratios are smaller, but they are still better than 40:1, with a maximum value of 70:1. The fabricated ASR PBS also has very high efficiencies. The measured reflection efficiency for the TE-polarized light and transmission efficiency for the TM-polarized light are higher than 99% and 97%, respectively. The slightly lower measured efficiency for the TM-polarized light may occur because of the etch depth errors, causing reductions of reflection polarization extinction ratios. We expect that more-accurate control of fabrication tolerances (i.e., etch depth) will improve both the efficiencies and the extinction ratios.

For comparison between the experimental results and the numerical design predictions, we used the SEM image (see Fig. 8) and the optical microscope observation of the fabricated element to estimate the fabrication tolerances of our sample. The sample observed under the optical microscope shows that the SiO_2 substrate areas have a light brown color, indicating that the last Si layer was underetched. The SEM photograph indicates a deviation of the grating profile from exact rectangular shape. For modeling purposes we estimate that our sample is 6 nm underetched. Consequently, we use the modified grating profile shown in Fig. 10(a). Figures 10(b) and 10(c) show that the experimental results and the numerical predictions are in good agreement. Note that the only significant deviation between modeling and experimental results occurs in Fig. 10(c) near the incidence angle of 46°. This deviation may occur because of the limited dynamic range of our photodetectors (i.e., a dynamic range of more than 10^5 will be needed to resolve this resonance).

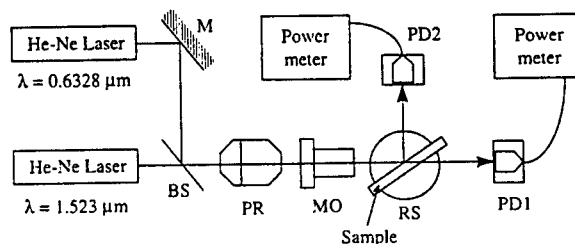


Fig. 9. Schematic diagram of the experimental setup for the characterization of the fabricated ASR PBS. M, mirror; PR, polarization rotator; BS, beam splitter; MO, microscope objective; RS, rotation stage; PD1 and PD2, Ge photodetectors. The transmittance and the reflectance are measured simultaneously to ensure accurate comparison of the extinction ratios and efficiencies.

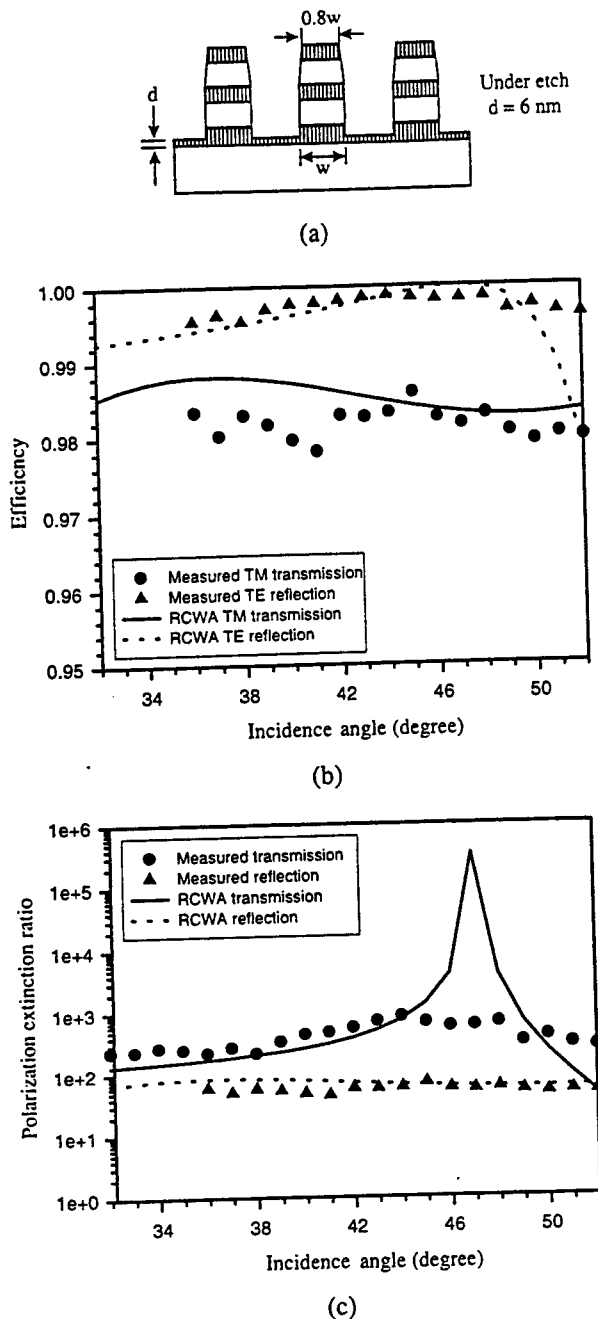


Fig. 10. Comparison of experimental measurements and numerical predictions of the fabricated ASR PBS. (a) The device structure shown in Fig. 3(a) has been modified to account for the fabricated grating profile error as well as the underetching fabrication error. (b) Measured and calculated efficiencies for TE- and TM-polarized waves, (c) measured and calculated polarization extinction ratios in transmission and reflection versus incidence angle.

6. CONCLUSIONS

We described the design, fabrication, and experimental evaluations of an ASR PBS. This novel element combines the form-birefringence effect of a high-spatial-frequency grating with the high reflectance of multilayer structures. We use EMT for initial design and RCWA for

optimization and numerical characterization of the PBS. Numerical modeling has also shown that these elements can tolerate broad-spectrum optical radiation with good performance in a wide 2D angular bandwidth range. We have demonstrated that our PBS can be designed to adapt easily to different applications, such as highly efficient polarization-selective mirrors for vertical-cavity microlaser and broadband low-insertion-loss normal-incidence polarizers. Other ASR PBS designs for specified spectrum ranges can be attained by means of suitable combinations of materials to construct the multilayer structures. This design flexibility shows how ASR PBS can address the needs of many different optoelectronic packaging systems. We also discussed the influence of different fabrication errors on the performance of the PBS and observed a rather large fabrication error tolerance of the device.

A sample of the ASR PBS designed for a wavelength of $1.523 \mu\text{m}$ has been fabricated and experimentally characterized. The experimental results show that our PBS provides high measured polarization extinction ratios (maximum value of 830:1) for the two orthogonally polarized output optical waves. The device has also been shown to have the capabilities for operating with optical signals of wide angular bandwidth ($\pm 10^\circ$ near the designated incidence angle of 42°) with a high extinction ratio ($>220:1$) and high efficiencies ($>97\%$). Finally, the experimental results have been compared with the numerical predictions and have been found in good agreement. Our future research goals are directed toward enhancing the fabrication precision and toward developing new ASR PBS designs with suitable materials for other spectrum range.

The ASR PBS combines such unique features as compactness, compatibility with semiconductor materials, negligible insertion losses, polarization selectivity for light at normal incidence, high polarization extinction ratios, and operation with waves of large angular bandwidth and from a broad spectral range. These characteristics make the devices desirable for use in image processing, in optical interconnections, and in many other polarization optics applications.

ACKNOWLEDGMENTS

The authors thank Paul Shames for helpful discussions and preparation of the manuscript. The research conducted at the University of California, San Diego is supported in part by the National Science Foundation, the U.S. Air Force Office of Scientific Research, and the Rome Laboratory. The Research conducted at Caltech is supported by the National Science Foundation.

REFERENCES

1. F. B. McCormick, F. A. P. Tooley, T. J. Cloonan, J. L. Brubaker, A. L. Lentine, R. L. Morrison, S. J. Hinterlong, M. J. Herron, S. L. Walker, and J. M. Sasian, "Experimental investigation of a free-space optical switching network by using symmetric self-electro-optic-effect devices," *Appl. Opt.* **31**, 5431-5446 (1992).
2. Q. W. Song, M. C. Lee, and P. J. Talbot, "Polarization sensitivity of birefringent photorefractive holograms and its

- applications to binary switching," *Appl. Opt.* **31**, 6240-6246 (1992).
3. Y.-T. Huang and Y.-H. Chen, "Polarization-selective elements with a substrate-mode grating pair structure," *Opt. Lett.* **18**, 921-923 (1993).
 4. M. Ojima, A. Saito, T. Kaku, M. Ito, Y. Tsunoda, S. Takayama, and Y. Sugita, "Compact magneto-optical disk for coded data storage," *Appl. Opt.* **25**, 483-489 (1986).
 5. R. K. Kostuk, T.-J. Kim, G. Campbell, and C. W. Han, "Diffractive-optic polarization-sensing element for magneto-optic storage heads," *Opt. Lett.* **19**, 1257-1259 (1994).
 6. P. Kunstmann and H.-J. Spitschan, "General complex amplitude addition in a polarization interferometer in the detection of pattern differences," *Opt. Commun.* **4**, 166-168 (1971).
 7. L. B. Wolff, "Polarization camera for computer vision with a beam splitter," *J. Opt. Soc. Am. A* **11**, 2935-2945 (1994).
 8. K. Shiraishi, T. Sato, and S. Kawakami, "Experimental verification of a form-birefringent polarization splitter," *Appl. Phys. Lett.* **58**, 211-212 (1991).
 9. J. L. Pezzaniti and R. A. Chipman, "Angular dependence of polarizing beam-splitter cubes," *Appl. Opt.* **33**, 1916-1929 (1994).
 10. S. Aoyama and T. Yamashita, "Grating beam splitter polarizer using multi-layer resist method," in *International Conference on the Application and Theory of Periodic Structures*, J. M. Lerner and W. R. McKinney, eds., *Proc. SPIE* **1545**, 241-250 (1994).
 11. M. C. Gupta and S. T. Peng, "Multifunction grating for signal detection of optical disk," in *Optical Data Storage '91*, T. A. Shull, N. Imamura, and J. J. Burke, eds., *Proc. SPIE* **1499**, 303-306 (1991).
 12. S. Habraken, O. Michaux, Y. Renotte, and Y. Lion, "Polarizing holographic beam splitter on a photoresist," *Opt. Lett.* **20**, 2348-2350 (1995).
 13. R. M. A. Azzam, "Polarizing beam splitters for infrared and millimeter waves using single-layer coated dielectric slab or unbacked films," *Appl. Opt.* **25**, 4225-4227 (1986).
 14. R. C. Tyan, P. C. Sun, A. Scherer, and Y. Fainman, "Polarizing beam splitter based on anisotropic spectral reflectivity characteristics of form-birefringent multilayer gratings," *Opt. Lett.* **21**, 761-763 (1996).
 15. S. M. Rytov, "Electromagnetic properties of a finely stratified medium," *Sov. Phys. JETP* **2**, 466-475 (1956).
 16. M. G. Moharam and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," *J. Opt. Soc. Am.* **72**, 1385-1392 (1982).
 17. I. Richter, P. C. Sun, F. Xu, and Y. Fainman, "Design considerations of form birefringent microstructures," *Appl. Opt.* **34**, 2421-2429 (1995).
 18. D. F. Edwards and H. R. Philipp, in *Handbook of Optical Constants of Solids*, E. D. Palik, ed. (Academic, Orlando, Fla., 1985), pp. 547 and 749.
 19. E. Hecht, *Optics* (Addison-Wesley, Reading, Mass., 1990), p. 377.
 20. E. D. Palik, in *Handbook of Optical Constants of Solids*, E. D. Palik, ed., (Academic, Orlando, Fla., 1985), p. 695.
 21. F. Xu, R.-C. Tyan, P.-C. Sun, Y. Fainman, C.-C. Cheng, and A. Scherer, "Fabrication, modeling, and characterization of form-birefringent nanostructures," *Opt. Lett.* **20**, 2457-2459 (1995).
 22. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1975), pp. 705-708.
 23. Reference 22, pp. 51-70.
 24. C. W. Haggans, L. Li, and R. K. Kostuk, "Effective-medium theory of zeroth-order lamellar gratings in conical mountings," *J. Opt. Soc. Am. A* **10**, 2217-2225 (1993).
 25. H. Kikuta, H. Yoshida, and K. Iwata, "Ability and limitation of effective medium theory for subwavelength gratings," *Opt. Rev.* **2**, 92-99 (1995).
 26. F. Xu, R.-C. Tyan, Y. Fainman, and J. E. Ford, "Single-substrate birefringent computer-generated holograms," *Opt. Lett.* **21**, 516-518 (1996).

A. 8. Fabrication, modeling, and characterization of form-birefringent nanostructures

F. Xu, R.C. Tyan, P.C. Sun and Y. Fainman

Optics Letters, Vol. 20, No. 24, December 15, 1995

Fabrication, modeling, and characterization of form-birefringent nanostructures

Fang Xu, Rong-Chung Tyan, Pang-Chen Sun, and Yeshayahu Fainman

Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92037

Chuan-Cheng Cheng and Axel Scherer

Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125

Received August 15, 1995

A 490-nm-deep nanostructure with a period of 200 nm was fabricated in a GaAs substrate by use of electron-beam lithography and dry-etching techniques. The form birefringence of this microstructure was studied numerically with rigorous coupled-wave analysis and compared with experimental measurements at a wavelength of 920 nm. The numerically predicted phase retardation of 163.3° was found to be in close agreement with the experimentally measured result of 162.5° , thereby verifying the validity of our numerical modeling. The fabricated microstructures show extremely large artificial anisotropy compared with that available in naturally birefringent materials and are useful for numerous polarization optics applications.

© 1995 Optical Society of America

The form-birefringence or artificial-birefringence effect occurs when the period of such microstructures is much less than the wavelength of the incident optical field and the far field of the transmitted radiation will possess only zero-order diffraction. The two prevalent approaches to characterize such artificial dielectric properties of the microstructured boundary use the effective medium theory¹ and rigorous coupled-wave analysis^{2,3} (RCWA). In this study we choose to use RCWA because the simpler effective medium theory does not provide accurate results when the microstructured grating period approaches the wavelength of the radiation.^{3,4} Form-birefringent nanostructures (FBN's) have several unique properties³ that make them superior to naturally birefringent materials: (i) A high value for the strength of form birefringence, $\Delta n/n$, can be obtained by the selection of substrate dielectric materials with a large refractive-index difference (here Δn and n are the difference and the average effective indices of refraction, respectively, for the two orthogonal polarizations); for example, a high-spatial-frequency surface-relief grating of rectangular profile on a GaAs substrate provides a $\Delta n/n$ value of ~ 0.63 , which is much larger than those found for naturally birefringent materials (e.g., for calcite the value of $\Delta n/n$ is ~ 0.1). (ii) The magnitude of form birefringence, Δn , can be adjusted by variation of the duty ratio as well as of the shape of the microstructures.³ (iii) FBN's can be used to modify the reflection properties of the dielectric boundaries.^{3,5} Such FBN's are useful for constructing polarization-selective beam splitters^{6,7} and general-purpose polarization-selective diffractive optical elements such as birefringent computer-generated holograms⁸ (BCGH's).

A BCGH is a general-purpose diffractive optical element that has two independent though arbitrary impulse responses for the two orthogonal linear polarizations. BCGH elements are useful in various appli-

cations.⁸ In its original design⁹ a BCGH consists of two surface-relief substrates with at least one of them birefringent. The two independent etch depths of the BCGH element provide the two degrees of freedom necessary to encode the two independent phase functions. However, the BCGH fabrication process can be simplified by use of a single FBN made of an isotropic substrate. One can obtain the two degrees of freedom necessary for construction of a BCGH by varying, for example, the duty ratio and the etch depth of the dielectric nanostructures. In this Letter we investigate the fabrication and characterization of FBN's to determine their usefulness for construction of a form-birefringent computer-generated hologram.

Fabrication of FBN's for visible and near-infrared wavelength regions is a challenging task. In the past, artificial birefringence was observed experimentally for microstructures with a relatively large period that can be operated in the microwave or far-infrared spectrum range.⁵ For visible and near-infrared radiation-range applications the artificial dielectrics were fabricated with a stratified multilayer structure¹⁰ or by the recording of interference patterns of two coherent light beams to create a subwavelength grating in a photoresist.¹¹ Neither method is suitable for our BCGH applications since the former creates the form birefringence in a direction perpendicular to the substrate surface and the latter does not provide design flexibility in terms of microstructure shape and the values of the dielectric constants. To achieve the design and the fabrication flexibility required by a BCGH, we use electron-beam lithography to generate the high-spatial-frequency patterns.

The fabrication procedures of FBN's in GaAs substrates are shown schematically in Fig. 1. First, a GaAs substrate was coated with a layer of SiO₂, then a layer of Au, and finally a layer of high-molecular-weight poly(methyl methacrylate) (PMMA). Electron-beam lithography with a 30-kV

incident beam energy was used to define the high-resolution linear gratings over a square area of $100\ \mu\text{m} \times 100\ \mu\text{m}$ on the spun-on 70-nm-thick resist layer. The PMMA pattern was developed for 14 s in a 3:7 mixture of $\text{C}_2\text{H}_5\text{OCH}_2\text{CH}_2\text{OH}$ in CH_3OH and then was transferred onto the 70-nm-thick Au layer by ion milling with 1500-V Ar ions. This Au layer was used as a dry-etching mask to transfer the patterns into the 100-nm-thick layer of sputter-deposited SiO_2 by reactive-ion etching. During this etching process, 60-mTorr C_2F_4 was used as the reactive gas, and a 300-V bias voltage was applied (50 W of rf power) at an etch rate of 20 nm/min. Then, a chemically assisted ion-beam etching system helped to etch the high-resolution nanostructure to the desired depth in the GaAs by using an Ar-ion beam assisted with Cl_2 reactive gas. Finally, we removed the SiO_2 mask by immersing the sample into buffered HF. The 490-nm-deep nanostructure with a period of 200 nm fabricated in GaAs substrate was inspected under a scanning-electron microscope (SEM). The top view and the cross-sectional view of the fabricated nanostructure are shown in Figs. 2(a) and 2(b), respectively.

The experimental setup for characterization of the form birefringence of fabricated nanostructures is shown schematically in Fig. 3. An Ar⁺-pumped Ti:sapphire laser was operated at a wavelength of 920 nm, where the GaAs substrate is transparent with minimum absorption. The polarization of the laser beam was controlled by a polarization rotator so that the normally incident optical wave was polarized linearly at 45° with respect to the grooves' direction. We used a microscope objective to focus the incident beam onto the $100\ \mu\text{m} \times 100\ \mu\text{m}$ microstructure pattern. At a distance of 1 m from the sample we inserted a 1-cm-diameter aperture stop and a polarization analyzer followed by a photodetector. The aperture stop was introduced to avoid contributions of the obliquely incident light and diffracted field from the edges of the sample, thereby ensuring the validity of the paraxial approximation necessary for our polarization measurements. A Glan-Thompson-type polarizer was used as an output analyzer.

For the experimental characterization of the FBN we used Jones calculus. Let the Jones matrix of the form-birefringent nanostructure on a GaAs substrate be given by

$$J_s = \begin{bmatrix} a & 0 \\ 0 & b \exp(j\phi_s) \end{bmatrix}, \quad (1)$$

where a and b are the amplitude transmittances of the horizontally and vertically polarized light (i.e., perpendicular and parallel to the grooves' direction), and ϕ_s is the phase difference between them on propagation through the FBN. The output fields for these two polarizations can be formulated as

$$\begin{bmatrix} E_{H\text{out}} \\ E_{V\text{out}} \end{bmatrix} = J(\theta) \begin{bmatrix} a & 0 \\ 0 & b \exp(j\phi_s) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{E_i}{\sqrt{2}}, \quad (2)$$

where

$$J(\theta) = \mathbf{R}(-\theta)\mathbf{J}(0)\mathbf{R}(\theta)$$

$$= \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (3)$$

represents the Jones matrix of an analyzer that is aligned at an angle θ with respect to the vertical direction, and the input field is 45° linearly polarized. The intensity measured at the detector is given by

$$I_{\text{out}} = (a^2 \cos^2 \theta + b^2 \sin^2 \theta + 2ab \sin 2\theta \cos \phi_s) \frac{I_{\text{in}}}{2}. \quad (4)$$

Figure 4 shows a typical curve of measured intensity versus the orientation angle of the analyzer, θ , in the setup of Fig. 3. The two curves correspond to the measurements of the GaAs substrate with and without the FBN. We curve fitted the measured data by using Eq. (4) (see the solid and dashed curves in Fig. 4), which yielded the resultant parameters $a = 0.67$, $b = 0.57$, and $\phi_s = 162.5^\circ$. Note

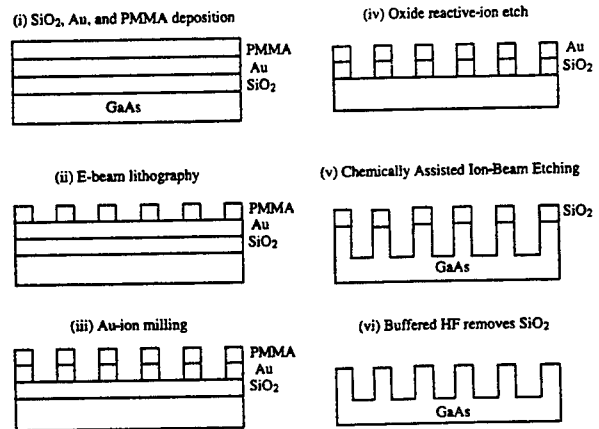


Fig. 1. Schematic of the procedures for the fabrication of form-birefringent nanostructures in GaAs substrates.



Fig. 2. SEM photographs of the fabricated form-birefringent nanostructure in a GaAs substrate: (a) top and (b) cross-sectional views.

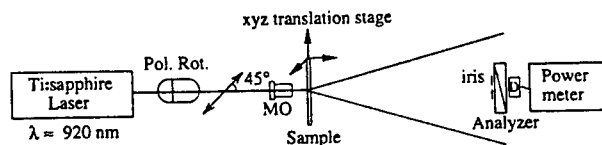


Fig. 3. Schematic of the experimental setup for the characterization of form-birefringent nanostructures. Pol. Rot., polarization rotator; MO, microscope objective.

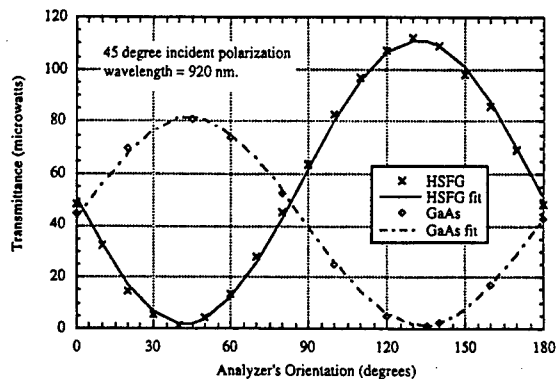


Fig. 4. Experimental measurements and curve-fitted results of the transmitted intensity versus the orientation of the analyzer for a GaAs substrate with and without the form-birefringent high-spatial-frequency grating (HSFG).

that ϕ_s is the positive phase difference between the fields at vertical and horizontal polarizations because the effective index for polarization parallel to the grooves of the nanostructure is larger than that for the perpendicular polarization.³ From Fig. 4 we can also observe that the transmittance from the GaAs substrate with the FBN is larger than that from the GaAs substrate alone. This effect is obtained because the effective indices of the nanostructure for both polarizations are smaller than those of the GaAs substrate, and therefore the nanostructure pattern acts as an antireflection coating.

The parameters a , b , and ϕ_s were also calculated numerically with RCWA applied to the measured profile [Fig. 2(b)] of the fabricated FBN in GaAs substrate. The profile is described by a trapezoid shape, with its top edge being 5% of the period and the bottom edge being 95% of its period. The period and the depth of the GaAs nanostructure were estimated from the SEM photographs of Fig. 2 to be 200 and 490 nm, respectively. The GaAs substrate is 500 μm thick, with a refractive index of 3.57 and an absorption coefficient of 3.25×10^{-5} , which are interpolated from the data in Ref. 12. The numerical simulations provide parameters $a = 0.743$, $b = 0.714$, and $\phi_s = 163.3^\circ$. The computer-simulation result for the phase difference between the two orthogonal polarizations is found to be in very good agreement (0.5% difference) with the measured results, confirming the validity of our RCWA-based numerical model. We anticipate that the slight difference in the amplitude transmission coefficients occurs as a result of (i) some scattering loss on the surface of the nanostructure, (ii) diffraction scattering on the limiting aperture of the nanostructure, (iii) inaccuracy in the calculated absorption coefficient for the GaAs substrate, and (iv) inaccuracy in the assumed profile and depth.

The characteristics of the fabricated nanostructure shown in Fig. 4 indicate that it will be possible to obtain a relative phase retardation of π (e.g., a half-wave plate) between the vertical and horizontal polari-

zations. Note that, by rotating the orientation of the periodic nanostructure on the GaAs substrate by 90° , we will obtain the negative value $-\pi$ for the phase retardation between the vertical and horizontal polarizations. Therefore by controlling the orientation of the periodic nanostructure we will be able to obtain a total range of phase retardation between $-\pi$ and π , which will be sufficient for the design of a binary-phase single-substrate BCGH. Furthermore, this phase-retardation range will be useful for encoding the phase difference of a multiple-phase-level BCGH, whereas absolute relative phase will need to be corrected by mean of other methods.

In conclusion, we have fabricated a 490-nm form-birefringent nanostructure with a period of 200 nm in a GaAs substrate. Form birefringence of the nanostructure was studied numerically with RCWA and compared with experimental measurements at a wavelength of 920 nm. The theoretical modeling used the grating profile measured from SEM photographs of these nanostructures. The predicted phase retardation of 163.3° is found to be in close agreement with the experimentally measured result of 162.5° . Controlling the orientation of the dielectric nanostructure permits us to obtain a phase retardation varying from $-\pi$ to π for the two orthogonal linear polarizations. The fabricated nanostructures show extremely large artificial anisotropy compared with that available in naturally birefringent materials and are useful not only for single-substrate form-birefringent computer-generated holograms³ but also for numerous other polarization optics applications.

This study was funded by the National Science Foundation, the U.S. Air Force Office for Scientific Research, the Optical Technology Center of the Advanced Research Projects Agency, and the Rome Laboratory.

References

1. S. M. Rytov, *Sov. Phys. JETP* **2**, 466 (1956).
2. E. N. Glytsis and T. K. Gaylord, *Appl. Opt.* **31**, 4459 (1992).
3. I. Richter, P. C. Sun, F. Xu, and Y. Fainman, *Appl. Opt.* **34**, 2421 (1995).
4. H. Kikuta, H. Yoshida, and K. Iwata, *Opt. Rev.* **2**, 92 (1995).
5. D. H. Raguin and G. M. Morris, *Appl. Opt.* **32**, 1154 (1993).
6. A. Ohba, Y. Kimura, S. Sugama, Y. Urino, and Y. Ono, *Jpn. J. Appl. Phys. Suppl.* **28**, 359 (1989).
7. P. Kipfer, M. Collischon, H. Haidner, J. Schwider, N. Streibl, and J. Lindolf, *Opt. Eng.* **33**, 79 (1994).
8. F. Xu, J. E. Ford, and Y. Fainman, *Appl. Opt.* **34**, 256 (1995).
9. J. E. Ford, F. Xu, K. Urquhart, and Y. Fainman, *Opt. Lett.* **18**, 456 (1993).
10. T. Sato, K. Shiraishi, K. Tsuchida, and S. Kawakami, *Appl. Phys. Lett.* **61**, 2633 (1992).
11. L. H. Cescato, E. Gluch, and N. Streibl, *Appl. Opt.* **29**, 3286 (1990).
12. E. D. Palik, ed., *Handbook of Optical Constants of Solids* (Academic, New York, 1985).

A. 9. Polarizing Beam Splitters Constructed of Form-Birefringent Multilayer Gratings

R.C. Tyan, P.C. Sun and Y. Fainman

SPIE Proceedings: Diffractive and Holographic Optics Technology, Vol.2689

San Jose, CA 1996.

Polarizing beam splitters constructed of form-birefringent multilayer gratings

Rong-Chung Tyan, Pang-Chen Sun and Yeshayahu Fainman

Department of Electrical and Computer Engineering, University of California at San Diego
La Jolla, CA 92093

ABSTRACT

We introduce a novel polarizing beam splitter that uses the anisotropic spectral reflectivity (ASR) characteristics of a high spatial frequency multilayer binary grating. By combining the form birefringence effect of a high spatial frequency grating with the resonant reflectivity of a periodic multilayer structure, the ASR characteristics for the two orthogonal linear polarizations are obtained. Such ASR effects allow us to design an optical element that is transparent for TM polarization but reflective for TE polarization. The properties of the polarizing beam splitter are investigated using rigorous coupled-wave analysis. The design results show that an ASR polarization beam splitter can provide a high polarization extinction ratio for optical waves from a wide range of incident angles and a broad optical spectral bandwidth. Such ASR polarizing beam splitters are uniquely suitable for image processing and optical interconnection applications.

Keywords : polarizing beam splitter, high spatial frequency binary grating, multilayer structure, form birefringence, diffractive optical element, optical components.

1. INTRODUCTION

Polarizing beam splitters (PBS) are essential components for numerous optical information processing applications such as free-space optical switching networks¹, read-write magneto-optic data storage systems², and polarization based imaging systems³. These applications require that the PBS providing high extinction ratios tolerate a wide angular bandwidth, a broad wavelength range of the incident waves, and compact size for efficient packaging. Conventional PBS employing either natural crystal birefringence (e.g., Wollaston prisms) or polarization selectivity of multilayer structures (e.g., PBS cubes) do not meet these requirements. The Wollaston prism requires a large thickness to generate enough walk-off distance between the two orthogonal polarizations due to intrinsically small birefringence of the naturally anisotropic materials. An alternative design of Wollaston prisms⁴ reduces the thickness considerably by taking advantages of periodic multilayer slab structures that possess form birefringence which is several times larger than that of natural birefringent materials. However, the fabrication of such a multilayer slab structure is a tedious and long process. PBS cubes are easier to fabricate, but they provide good extinction ratios only in a narrow angular bandwidth for a limited wavelength range⁵. Other designs which utilize form birefringent high spatial frequency surface relief gratings⁶ and single-layer-coated dielectric slab⁷, have also been proposed to reduce the size of the components and to simplify the fabrication process, however, they also suffer from low extinction ratio, small operating angular bandwidth, and limited wavelength range.

In this manuscript, we introduce a new PBS device that uses the unique properties of anisotropic spectral reflectivity (ASR) characteristics of a high spatial frequency multilayer binary grating. The new ASR mechanism is based on combining the effect of form birefringence of a high spatial frequency grating (i.e., grating period is much less than the wavelength of the incident field) with the resonant reflectivity of a multilayer structure. In the next section we first describe intuitively the principle of the ASR behavior of the high spatial frequency multilayer grating using Effective Medium Theory (EMT)⁸. Then we use Rigorous Couple-Wave Analysis (RCWA)^{9,10} tools for an optimum design of the PBS, where the EMT results are used as an initial estimate. In sections 3 and 4 respectively, we use RCWA to design the ASR polarizing beam splitters and characterize them in terms of polarization extinction ratios for operation with waves of wide angular bandwidth and broad wavelength range. The results demonstrate extremely high extinction ratios (e.g., 1000000:1) when the PBS is operated at a specified wavelength and angle of incidence. Furthermore, good average extinction ratios (from 800:1 to 50:1) can be obtained when the PBS is operated for waves of 20° angular bandwidth with wavelength ranging from 1300nm to 1500nm. The conclusions and future research directions are discussed in section 5.

2. ANISOTROPIC SPECTRAL REFLECTIVITY

Consider a multilayer structure formed on a substrate by depositing alternating layers of dielectric materials with high and low indices of refraction, n_h and n_l respectively. Such a structure exhibits high reflectivity in a wide spectral bandwidth, particularly when the thickness of each layer corresponds to a quarter-wave optical thickness for the center wavelength. The reflectivity of the quarter-wave structure can be increased by increasing the value of the ratio n_h/n_l and the number of layers in the stack. Larger values of n_h/n_l also increase the spectral bandwidth of high reflectance. For a multilayer structure made of isotropic dielectric materials, the reflectivity spectrums for the two orthogonal linear polarizations are identical and therefore, hardly separable. To separate the reflectivity spectrums for the two orthogonal linear polarizations, we need to substitute one or both (i.e., high and low refractive indices) materials with birefringent materials. Such a multilayer structure of anisotropic materials will possess reflectivity spectrum bands centered at different wavelengths for the two orthogonal polarizations, thereby, providing the desired separation of reflectivity spectrums. However, since natural materials possess very small birefringence, the separation of the reflection spectral bands corresponding to the two orthogonal polarizations will be very limited. With our approach the separation of the reflection spectral bands for the two orthogonal linear polarizations can be considerably increased due to the high anisotropy¹¹ that can be obtained with form birefringence.

Form birefringence effects¹² appear in high spatial frequency gratings formed by isotropic dielectric materials. Due to the geometric anisotropy of the grating structure, the two orthogonally polarized optical fields, one parallel to the grating grooves (designated as TE field) and the other perpendicular to the grating grooves (designated as TM field), encounter different effective dielectric constants and thus acquire a phase difference between them. This is similar to that obtained in natural anisotropic materials. The magnitude of form birefringence depends on the geometric composition of the grating structure (including the dielectric indices and the shape of the grating)¹⁰ as well as the angle between the incident field and the grating vector. It is important to note that the value of form birefringence is a few times larger than the birefringence obtained with naturally birefringent materials. This makes the high spatial frequency grating an excellent candidate for separating the reflectivity spectrums for the two orthogonal polarizations.

Under the normal incidence, the effective indices for the TE and TM polarizations of a surface-relief high spatial frequency binary grating can be estimated from the 2nd order EMT⁸ :

$$n_{TE}^{(2)} = \left\{ n_{TE}^{(0)2} + \frac{1}{3} \left(\frac{\Lambda}{\lambda} \right)^2 \pi^2 F^2 (1-F)^2 (n_{III}^2 - n_I^2)^2 \right\}^{1/2} \quad (1)$$

$$n_{TM}^{(2)} = \left\{ n_{TM}^{(0)2} + \frac{1}{3} \left(\frac{\Lambda}{\lambda} \right)^2 \pi^2 F^2 (1-F)^2 \left(\frac{1}{n_{III}^2} - \frac{1}{n_I^2} \right)^2 n_{TE}^{(0)2} n_{TM}^{(0)6} \right\}^{1/2} \quad (2)$$

where F is the duty cycle of the grating defined by $F=1-a/\Lambda$ with a being the width of air gap in the grating (see Fig.1), Λ is the grating period, λ is the wavelength of the incident wave, n_I and n_{III} are the indices of air and the grating material, respectively, and $n_{TE}^{(0)} = [Fn_{III}^2 + (1-F)n_I^2]^{1/2}$ and $n_{TM}^{(0)} = \{n_{III}^2 n_I^2 / [Fn_I^2 + (1-F)n_{III}^2]\}^{1/2}$ are the effective indices of refraction for TE and TM waves provided by the zero order EMT.

Figure 1a shows an example of a high spatial frequency multilayer binary grating. We use SiO_2 and Si, with refractive indices of 1.45¹³ and 3.51¹⁴ respectively (for a wavelength of 1.3 μm) as the two materials for the multilayer structures because of their fabrication compatibility and low absorption coefficients in the near infrared region (this results in a low insertion loss). For operation of the form birefringent grating in the zero diffraction order we set the grating period equal to 0.5 μm and the duty cycle $F = 0.5$. Using second order EMT (eqs.1 and 2) we obtain the following effective refractive indices for the two materials $n_{TE,Si}^{(2)} = 3.25$, $n_{TE,SiO_2}^{(2)} = 1.26$ and $n_{TM,Si}^{(2)} = 1.71$, $n_{TM,SiO_2}^{(2)} = 1.18$. The effective indices of both materials are larger for TE polarization than for TM polarization. This means that in the spectral domain, the reflection band for TE polarization will be centered at a longer wavelength as compared with that for TM polarization. This ASR characteristic is the essential property needed to realize the PBS.

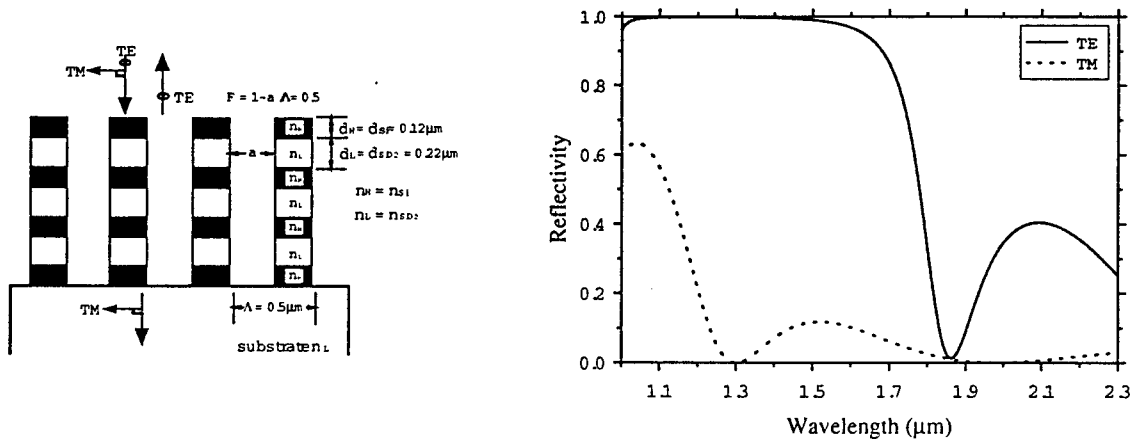


Fig. 1. (a) Schematic diagram of an ASR polarizing beam splitter operated with plane waves at normal incidence. (b) Numeric results of the reflectivity for TE and TM polarized waves vs. wavelength of a 7-layer PBS designed for normally incident waves.

Another characteristic is that the value of the effective index ratio for TE polarized light ($(n_h/n_l)_{TE}=2.58$) is larger than for TM polarized light ($(n_h/n_l)_{TM}=1.45$). This indicates that to achieve the same reflectance, the number of layers required by TE polarization will be less than that required by TM polarization. To minimize the number of layers needed to achieve a desired performance, we choose to maximize reflectivity for TE polarized light. Therefore, each layer has a quarter-wave optical thickness based upon the TE effective index. These values, estimated by EMT, are used as the basis for a more accurate design using RCWA. Optimization is done by incrementally varying the thickness of the layers to obtain the highest extinction ratio at the operational wavelength of $1.3\ \mu\text{m}$. To achieve broad reflectance peaks in the spectrum, we use high refractive index materials for both the first and the last layers in the structure. Figure 1b shows the numeric results of TE and TM reflectances as a function of the wavelength for a 7-layer high spatial frequency binary grating for normally incident optical fields. As expected, the reflectance peaks for TE and TM polarized light are separated and the TE polarization has a higher reflectance and broader bandwidth at longer wavelengths than the TM polarization. For the design wavelength of $1.3\ \mu\text{m}$, the TM polarized light will be transmitted while the TE polarized light will experience high reflectivity from the grating. The curves also show that the polarization extinction ratio remains high over a wide spectral range for the TM polarization. This feature allows the element to function as a low insertion loss polarizer. In fact, the sidelobe of the TM reflection is the only limit for the design of the polarization beam splitter. We anticipate that amplitude of the sidelobes may be reduced by fine-tuning other design parameters, such as the grating duty cycle. In the short wavelength region, both curves become irregular due to the coupling between the zero and the higher diffractive orders.

3. POLARIZING BEAM SPLITTER DESIGN

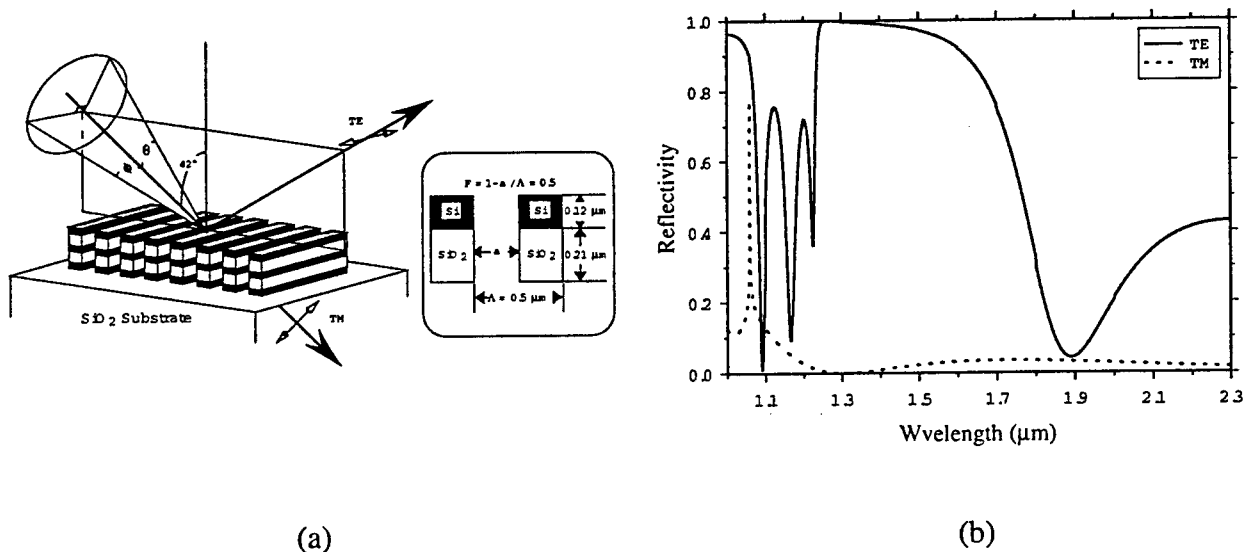


Fig. 2. (a) Schematic diagram of a 5-layer ASR polarizing beam splitter operated with incident waves at an angle of 42° . (b) Numeric results for the reflectivity of TE and TM polarized waves vs. wavelength for 42° incidence.

To realize a useful PBS that will allow us to separate the path of the reflected wave (i.e., TE polarized wave in our design) from that of the incident wave, we investigate a PBS design that operates with waves at large angles of incidence. Consider a geometry shown in Fig. 2a, where the input wave vector is introduced at 42° angle of incidence, lying in the plane perpendicular to the grating grooves and parallel to the grating vector. This slanted incidence arrangement possess two additional advantages: 1) reflectivity from each layer for TE polarization is increased, thus only five layers were needed to achieve the desired performance (normal incidence required seven); 2) the sidelobe for the TM reflectivity is flattened, allowing operation of the beam splitter in a wider spectral range. Here again we used first the EMT estimates for an accurate RCWA design. The thickness of each layer was first chosen to be a quarter-wave of the wavelength for TE polarized wave, and then fine-tuned to set the minimum of the TM reflectivity at the desired operating wavelength under the maximum band for the TE reflectivity. Since the reflection band for TE polarization is very wide, changing the thickness mostly affects the reflection band for TM polarization, and thus fine-tuning to achieve a desired performance is possible. Fig. 2b shows the numeric results of the reflectance vs. wavelength of the slanted incidence optical wave from the 5-layer grating. For the incident wavelength of $1.3\mu\text{m}$, the TE and TM reflectances are 0.9971 and 0.0009128, and the polarization extinction ratio for reflection side of the beam splitter is better than 1100:1. The highest extinction ratio of 1000000:1 is obtained at the wavelength of $\lambda = 1.265\mu\text{m}$. A relatively flat minimum zone of the TM reflectivity under the broadband of the TE reflectivity peak indicates that broadband operation is possible.

4. CHARACTERIZATION OF THE PBS PERFORMANCE

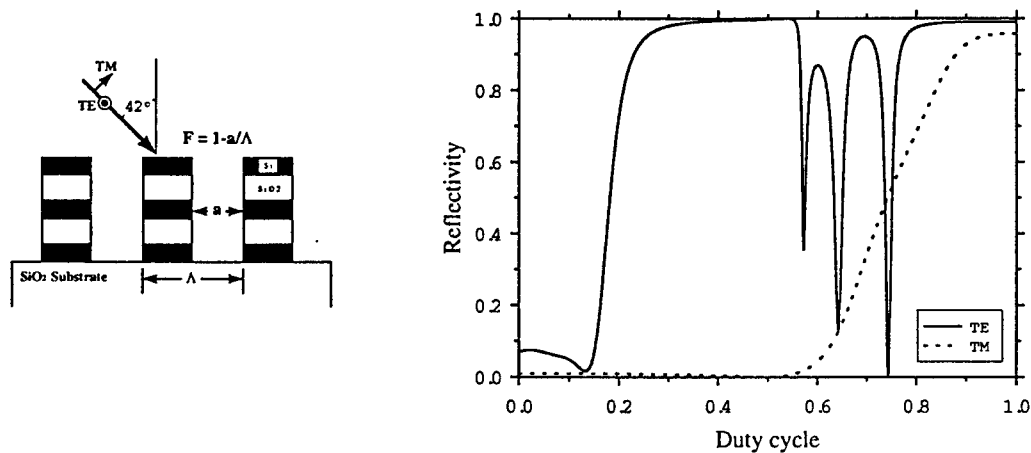


Fig. 3. Numeric results for the reflectivity of TE and TM polarized waves vs. duty cycle of the high spatial frequency binary grating. The design parameters are the same as using in Fig. 2a except the incident wavelength λ is now fixed at $1.3\mu\text{m}$ and the duty cycle F is varying from 0 to 1.

In order to find the tolerance of the ASR polarizing beam splitter to possible fabrication errors, we numerically characterize the performance of the PBS by changing the duty cycle of the multilayer gratings. We use the same design parameters of the PBS shown in Fig. 2a, except that the incident wavelength λ is fixed at $1.3\mu\text{m}$ and the duty cycle F is varied in the range 0 to 1.

The numeric results indicate that the reflectance of both polarizations stay approximately the same for the different duty cycles ranging from 0.3 to 0.55 (see Fig. 3). This shows high tolerance to fabrication errors of the ASR polarizing beam splitter. The reflectance of both polarizations become closer when the duty cycle approaches 0 and 1 and they become identical when the waves propagate at normal incidence.

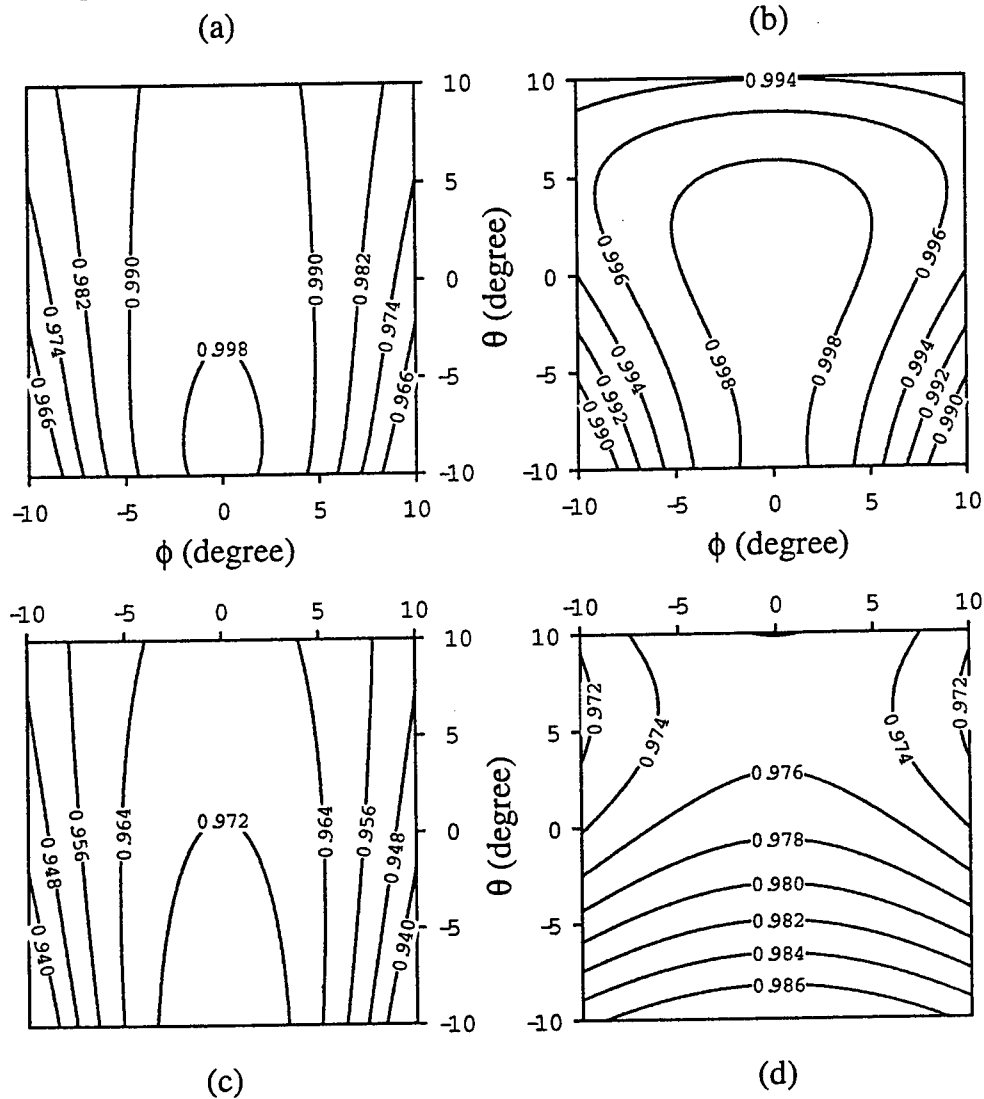


Fig. 4. Contour plots of TE reflectance and TM transmittance vs. incident angles (ϕ, θ) as defined in Fig. 2. (a) TE reflectance at wavelength $\lambda=1.3 \mu\text{m}$, (b) TM transmittance at wavelength $\lambda=1.3 \mu\text{m}$, (c) TE reflectance at wavelength $\lambda=1.5 \mu\text{m}$, and (d) TM transmittance at wavelength $\lambda=1.5 \mu\text{m}$.

We also investigate the angular dependence of the ASR polarizing beam splitter. As shown in Fig. 2a, the angles of incidence are varied to span an angular bandwidth of $\pm 10^\circ$ in both θ and ϕ directions defined around the initial 42° bias angle. The results shown in Fig. 4 indicate that, at wavelength $1.3 \mu\text{m}$, the reflectance for TE polarized light and the transmittance for TM polarized light are both better than 99% inside the 5° angular bandwidth cone, and better than 97% inside the 10° angular bandwidth cone. Around $1.5 \mu\text{m}$ results show that the TE reflectance

and TM transmittance from this PBS are still better than 96% inside the 5° angular bandwidth cone and better than 94% inside the 10° angular bandwidth cone. These results indicate that wide angular bandwidth, as well as broad spectral range, of operation is possible using this design. The performance of the PBS can be further improved by adding more layers to increase the range of high TE reflectivity as well as bring the efficiency closer to 100%.

Fabrication of such PBS in the visible spectral range is a challenging task due to the need to fabricate a grating with sub-wavelength grating period. However, for near infrared range of operation, fabrication of the structure is practical. For example, in our design the total grating depth of 5 layers is only $0.78\mu\text{m}$ with the grating period $0.5\mu\text{m}$, resulting in a grating aspect ratio of about 3:1, which is well within the fabrication capabilities of silicon microfabrication technology. Current thin film coating technology allows us to control the accuracy of the layer thickness within a few nanometers. Therefore, fabrication of the designed PBS (shown in Fig. 2a) can be done by first fabricating the multilayer structure, followed by direct e-beam lithography and ion beam etching a binary grating profile.

5. CONCLUSION

In conclusion, we have introduced a novel PBS device that is based on the ASR characteristics of a high spatial frequency multilayer binary grating. This PBS combines the form birefringence effect of a high spatial frequency grating with the high reflectance of multilayer structures. We use EMT for initial design and RCWA for optimization of the ASR polarizing beam splitter. We numerically characterize the ASR polarizing beam splitter in terms of polarization extinction ratio for operation with waves of wide angular bandwidth and broad wavelength range. The results show that the ASR polarizing beam splitter not only provides a very high extinction ratio for the two orthogonal polarizations, but also can be operated with optical signals of wide angular bandwidth and broad spectral range. From the numeric results, the fabrication error tolerance of the PBS has been shown to be very high. Another important advantage of the ASR polarizing beam splitters are their negligible insertion losses achieved by using non-absorbing dielectric materials. The ASR polarizing beam splitters combine such unique features as small size, negligible insertion losses, high polarization extinction ratios, and operation with waves of large angular bandwidth and broad spectral range. These features make these devices desirable for use in optical image processing, optical interconnections as well as other polarization optics applications.

6. ACKNOWLEDGMENTS

Authors thanks Paul Shames and Fang Xu for helpful discussions and preparation of the manuscript. This work was supported in part by the National Science Foundation, Advanced Research Projects Agency, Air Force Office Scientific Research and Rome Laboratory.

7. REFERENCES

1. F. B. McCormick, F. A. P. Tooley, T. J. Cloonan, J. L. Brubaker, A. L. Lentine, R. L. Morrison, S. J. Hinterlong, M. J. Herron, S. L. Walker, and J. M. Sasian, "Experimental investigation of a free-space optical switching network by using symmetric self-electro-optic-effect devices," *Appl. Opt.* Vol.31, pp.5431-5446 (1992).

2. M. Ojima, A. Saito, T. Kaku, M. Ito, Y. Tsunoda, S. Takayama, and Y. Sugita, "Compact magneto-optical disk for coded data storage," *Appl. Opt.* Vol.25, pp.483-489 (1986).
3. P. Kunstmann and H.-J. Spitschan, "General complex amplitude addition in a polarization interferometer in the detection of pattern differences," *Opt. Commun.* Vol.4, pp.166-168 (1971).
4. K. Shiraishi, and S. Kawakami, "Spatial walk-off polarizer utilizing artificial anisotropic dielectrics," *Opt. Lett.* Vol.15, pp.516-518 (1990).
5. J. L. Pezzaniti, and R. A. Chipman, "Angular dependence of polarizing beam-splitter cubes," *Appl. Opt.* Vol.33, pp.1916-1929 (1994).
6. M. C. Gupta and S. T. Peng, Proc. "Multifunction grating for signal detection of optical disk," *Optical Data Storage*, J. J. Burke, T. A. Shull, N. Imamura, ed. Vol.1499, pp.303-306, Proceedings of the SPIE - The International Society for Optical Engineering, (1991).
7. R. M. A. Azzam, "Polarizing beam splitters for infrared and millimeter waves using single-layer-coated dielectric slab or unbacked films," *Appl. Opt.* Vol.25, 4225 (1986).
8. S. M. Rytov, "Electromagnetic properties of a finely stratified medium," *Soviet Physics JETP*, Vol.2, pp.466-475 (1956).
9. M. G. Moharam, and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," *J. Opt. Soc. AM.* Vol.72, pp.1385-1392 (1982).
10. I. Richter, P. C. Sun, F. Xu, and Y. Fainman, "Design considerations of form birefringent microstructures," *Appl. Opt.*, Vol.34, pp.2421-2429 (1995).
11. F. Xu, R.-C. Tyan, P.-C. Sun, Y. Fainman, C.-C. Cheng, and A. Scherer, "Fabrication, Modeling and Characterization of Form Birefringent Nanostructures," *Opt. Lett.* (Accepted for publication in Dec. 1995).
12. M. Born, and E. Wolf, *Principles of Optics*, pp.705-708, Pergamon, Oxford, 1975.
13. H. R. Philipp, "Silicon dioxide (SiO₂) (glass)," *Handbook of Optical Constants of Solids*, E. D. Palik, ed., Vol.1, pp.749-763, Academic, Orlando, Fla., 1985.
14. D. F. Edwards, "Silicon (Si)," *Handbook of Optical Constants of Solids*, E. D. Palik, ed., Vol.1, pp.547-569, Academic, Orlando, Fla., 1985.

A. 10. Modeling and Optimization of Electro-Optic Phase Modulator

P. Shames, P.C. Sun and Y. Fainman

SPIE Proceedings: Physics and Simulation of Optical Devices IV, Vol.2693, pp. 787-96 San Jose, CA 1996.

Modeling and optimization of electro-optic phase modulator

P. Shames, P.C. Sun, and Y. Fainman

University of California, San Diego, Electrical and Computer Engineering Dept.
9500 Gilman Dr., La Jolla, California 92093

ABSTRACT

We introduce a novel method of modeling PLZT phase modulators. Traditionally, modeling has been based upon fitting the constant quadratic electro-optic coefficient to empirical data. Our characterization has shown that the electro-optic coefficient is not a constant and that the electro-optic effect saturates at electric field strengths that exist in standard surface electrode device configurations. We have also found that the additional effects of light scattering and depolarization, which depend on the strength of the applied electric field, are significant factors for modeling device design and optimization.

Keywords: PLZT, electro-optic effect, phase modulators, finite element analysis, depolarization, scattering.

1. INTRODUCTION

PLZT is an excellent material choice for use in spatial light modulators (SLM) due to its large electro-optic effect and low absorption for thin wafers¹. PLZT ceramics are used as transverse electro-optic modulators where the electric field is applied using interdigital surface electrodes (ISE). Such electro-optic devices are modeled based upon the quadratic electro-optic effect²⁻⁶ as well as its combination with the linear electro-optic effect⁷. However these models do not accurately predict the performance of an ISE device fabricated at UCSD⁸. In order to more accurately model such devices we experimentally characterized PLZT's electro-optic properties. Then we used finite element analysis (FEA) to characterize the field distributions for ISE devices. Finally, by combining the electric field values provided by FEA with the experimental electro-optic data, we were able to predict the performance of our ISE device. Although this methodology was carried out with PLZT 9.0/65/35 material, it can be applied to modeling electro-optic devices with arbitrary choice of material, electrode structure and geometry.

In the following section we will review the basic theory of quadratic electro-optic materials, as well as formulate the methodology for characterization of PLZT electro-optic material. In section 3, we will apply FEA to model the electric field induced by an electric potential difference between electrodes of ISE devices. The results of section 2 are integrated into the FEA model to determine the relationship between the change in relative phase of two orthogonally polarized components of an incident beam and the externally applied electric field. In section 4 we compare this prediction with the actual values taken from the fabricated ISE device. Conclusions and future directions are discussed in section 5.

2. CHARACTERIZATION OF ELECTRO-OPTIC MATERIAL

G. Haertling and C. Land initiated extensive studies characterizing PLZT ceramics¹. They found that thin wafers with compositions containing greater than 8 at.% La had strong electro-optic properties with transmission values of close to 100%. At room temperature, PLZT is isotropic due to its cubic crystallographic structure. When an external electric field is applied the PLZT material becomes polarized, demonstrating anisotropic optical characteristics. This same behavior is seen in crystals such as BaTiO₃ (in its cubic form) that exhibit primarily third order non-linear optical properties which in turn lead to quadratic electro-optic effects. Assuming PLZT follows this uniaxial crystal model, the optic axis will be determined by the direction of an externally applied electric field. The induced ordinary and extraordinary index of refraction is determined by

$$n_o = n - \frac{1}{2}n^3R_{12}E^2 \quad \text{and} \quad n_e = n - \frac{1}{2}n^3R_{11}E^2 \quad (1)$$

$$\Delta n(E) = n_o - n_e = \frac{1}{2}n^3RE^2 \quad (2)$$

where $R \equiv R_{12} - R_{11}$, R_{12} and R_{11} are the quadratic electro-optic coefficients, n is the refractive index of PLZT, and $\Delta n(E)$ is the induced optical birefringence. For PLZT with 9.0 at.% La the accepted value¹³ for R is approximately $3.8 \times 10^{16} \text{ m}^2/\text{V}^2$. Haertling and Land¹ noticed that the induced optical birefringence saturates with an externally applied electric field reaching a maximum⁷ value of 1.1×10^{-2} . More recently, M. Title⁵ mentioned the saturation effect in modeling embedded electrode PLZT modulators.

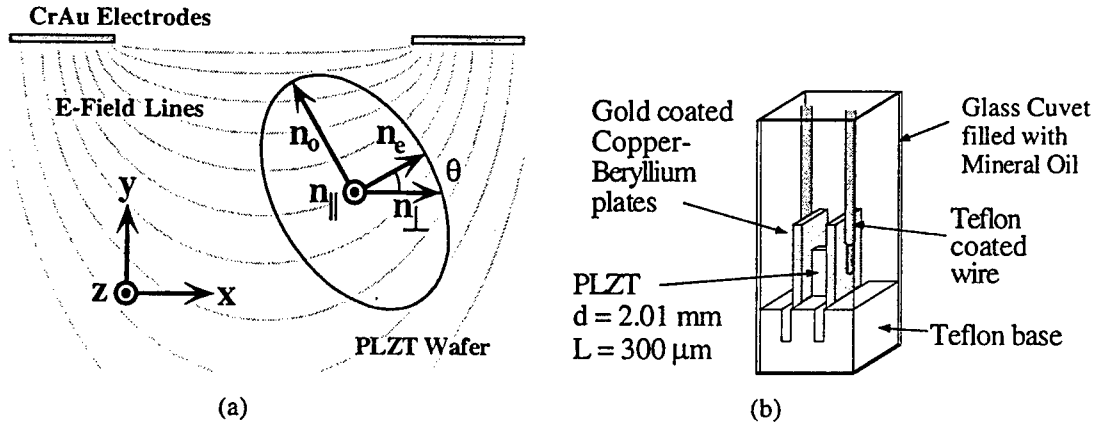


Figure 1. (a) For ISE on PLZT the index ellipsoid follows the tangent of the curved E-field lines. (b) PLZT wafer, 300 μm thick and 2.01 mm wide, placed between gold coated copper plates. The plates were inserted into a Teflon base to ensure good electrical isolation and the cuvet was filled with mineral oil to prevent current arcing due to exceeding the breakdown voltage of air.

A typical ISE device is constructed of stripe-shaped metal electrodes of width d and length L . An applied voltage across such electrodes creates curved lines of electric field within the PLZT (see Figure 1a). Since the optic axis follows the direction of the applied electric field, the axis orientation will vary as a function of position within the PLZT. Assuming the electrode length

$L \gg d$, the index of refraction parallel and perpendicular to the surface electrodes can be approximated⁵ by

$$n_{\parallel} = n_o \text{ and } n_{\perp} = \left[\frac{\cos^2(\theta(x, y))}{n_e^2} + \frac{\sin^2(\theta(x, y))}{n_o^2} \right]^{-\frac{1}{2}} \quad (3)$$

$$\text{where } \theta(x, y) = \tan^{-1} \left(\frac{E_y(x, y)}{E_x(x, y)} \right) \quad (4)$$

Defining the relative change in index $\Delta n(\theta(x, y)) = n_{\perp} - n_{\parallel}$, we determine the relative phase between the parallel and perpendicular polarization components of an optical beam passing through the device by

$$\Phi(x) = \frac{2\pi}{\lambda} \int_{y=0}^l \Delta n(\theta(x, y)) dy \quad (5)$$

where λ is the wavelength in vacuum and l is the thickness of the electrooptic material.

In the following we will experimentally determine the phase relationship provided by Equation 5. By placing the PLZT wafer between large parallel metal plates (see Figure 1b) a homogeneous electric field, perpendicular to the plates (i.e. $\theta = 0$), will exist within the dielectric. The field strength is a function of the potential difference between the plates (i.e. $E = \frac{V}{d}$, where d is the distance between the plates). To measure the induced phase retardation we illuminate the PLZT sample using a normally incident HeNe laser beam. The incident beam is linearly polarized at 45° with respect to the direction of the applied electric field, providing two equal components parallel and perpendicular with respect to the electrode structure. As the voltage, and thus the electric field strength, is varied, there will be a change in the relative phase between these two components.

By placing a crossed polarizer at the output of an ideal phase modulating device, the light intensity will vary as a function of the relative phase according⁹ to the relation

$$T = \frac{1}{2}a^2 + \frac{1}{2}b^2 - ab \sin(\Phi) \quad (6)$$

where Φ is the relative phase, and a^2 and b^2 are the transmittances for the two orthogonal components of the light. We measure the transmittance, T , through a crossed polarizer, as well as the transmittances through vertically and horizontally oriented polarizers, a^2 and b^2 respectively, as functions of the electric field. Then, by solving Equation 6 for Φ , we expect to find the relative phase as a function of electric field. However, using this method, the resultant phase is *not* continuous when a PLZT based device is being investigated. This is due to depolarization effects observed in PLZT phase modulators¹⁰.

By introducing a depolarizing term into the transmittance of the orthogonal components in Equation 6, we obtain

$$T = \frac{1}{2}(a^2 + c_a^2) + \frac{1}{2}(b^2 + c_b^2) - ab \sin(\Phi) \quad (7)$$

where c_a and c_b are the fractions of depolarized light corresponding to the incident vertically and horizontally polarized components. Defining $A^2 = a^2 + c_a^2$ and $B^2 = b^2 + c_b^2$, we get the relationship

$$T = \frac{1}{2}A^2 + \frac{1}{2}B^2 - \sqrt{A^2 - c^2}\sqrt{B^2 - c^2} \sin(\Phi) \quad (8)$$

where we also assume that $c_a \approx c_b = c$. Using the measured values for T, A^2, B^2 and curve fit for c^2 , we solve Equation 8 that provides continuous relative phase (see Figure 2b).

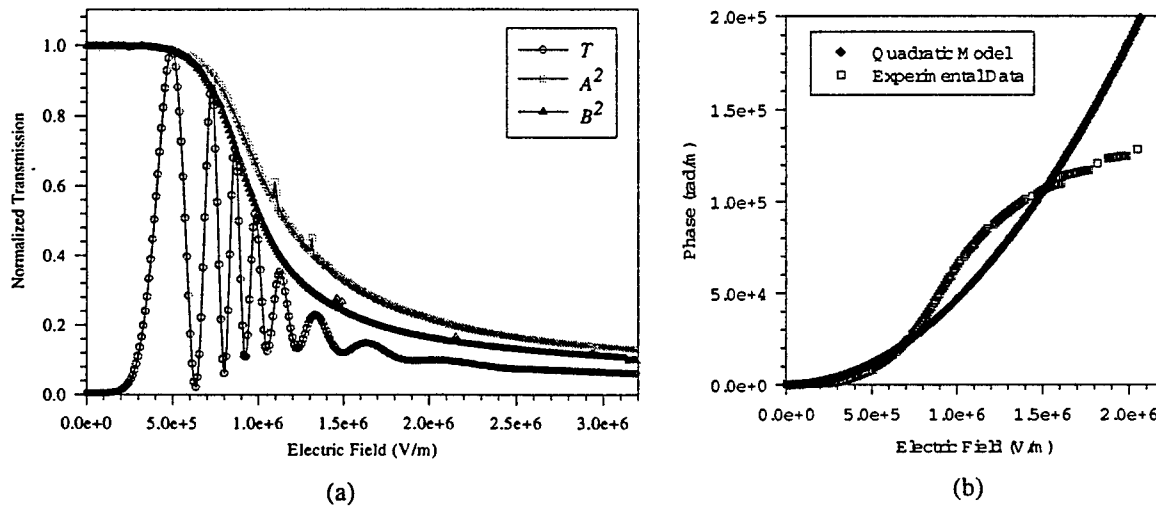


Figure 2. (a) Transmittance T through crossed polarizers set at 45° and through parallel polarizers set vertically and horizontally, A^2 and B^2 , as functions of a horizontally applied electric field. (b) Relative phase as a function of applied electric field. We also show the best fit for a quadratic electro-optic coefficient as $R = 6 \times 10^{-16} \left(\frac{\text{V}}{\text{m}}\right)^2$.

Observing the two curves A^2 and B^2 of Figure 2a we notice a dramatic intensity drop after the electric field reaches approximately 7×10^5 V/m. The main reason for the attenuation is due to scattering effects. This also causes a corresponding drop in the transmittance, T , through the crossed polarizers. Furthermore, T varies sinusoidally with the period varying as a function of electric field. The frequency of the sine initially increases, whereas at fields above 1×10^6 V/m the frequency decreases. Above the value of 2×10^6 V/m the contrast ratio approaches 1:1, which is primarily due to the depolarization effects. Due to the need for high contrast ratios in phase modulators, scattering and depolarization effects must be factored into the design of such devices.

Figure 2b shows the relative phase vs. electric field, where the experimental data is determined by solving Equation 8. Below electric fields of 2×10^5 V/m there is practically no phase change in the PLZT. Above values of 1×10^6 V/m we observe that the phase change begins to saturate. Our calculations show that for PLZT surface electrode devices, there are regions where the field strengths are on the order of 2×10^6 V/m, and therefore, the saturation effect must also be taken into consideration in these devices.

3. FINITE ELEMENT ANALYSIS OF ELECTRO-OPTIC MODULATORS

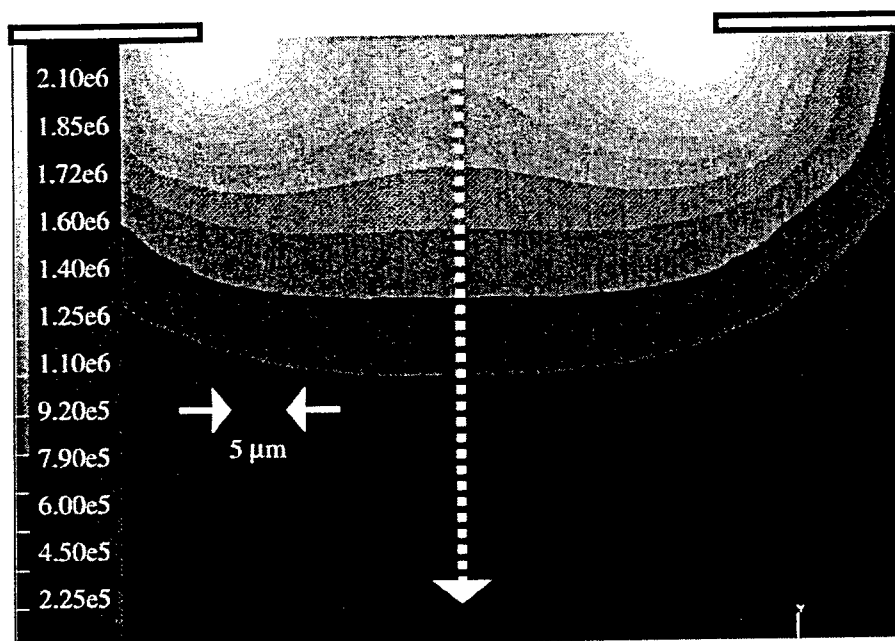


Figure 3. Mentat's output provides added insight into the effects of the electric field distribution. Shown is a portion of the contour band plot of the magnitude of the electric field between surface electrodes (outlined) $250 \mu\text{m}$ wide with a gap of $50 \mu\text{m}$. For field strengths below 2.25×10^5 V/m there is no phase modulation and above 2.0×10^6 V/m the modulation has reached a maximum (see Figure 2b). Therefore, modulation only occurs within $100 \mu\text{m}$ of the surface and near the edges of the electrodes we observe electric field strengths beyond the 'saturation' level.

Finite Element Analysis (FEA) is one of several methods^{3-6,11,12} available for calculating the electric field induced by metal electrodes. We use Mentat, a commercial FEA program from Marc Analysis, which provides an excellent tool for mesh generation, field calculations and visualization (see Figure 3). We used FEA to determine the electric field distribution in PLZT devices with surface electrodes. A typical example of Mentat's output is shown in Figure 3. For this particular ISE configuration, we observe that $100 \mu\text{m}$ from the surface the electric field strength drops below the minima required for phase modulation. We also see that near the edges of the electrodes the field goes beyond the maxima, i.e. the phase modulation has become constant. Using FEA we are able to calculate the electric field strength and direction, at any point, for any configuration of electrodes.

To find an explicit relationship between the magnitude and the direction of the electric field and the relative phase we substitute Eqs. 1, 2 and 3 into Equation 5 and obtain

$$\Phi(x) = \frac{2\pi}{\lambda} \int_{y=0}^l \left\{ n_o \left[\cos^2(\theta(x, y)) \left(1 - \frac{\Delta n(\theta=0)}{n_o} \right)^{-2} + \sin^2(\theta(x, y)) \right]^{\frac{1}{2}} - n_o \right\} dy \quad (9)$$

Using Taylor series expansion this can be approximated by

$$\begin{aligned} \Phi(x) &\approx \frac{2\pi}{\lambda} \int_{y=0}^l \left\{ n_o \left[1 + 2 \frac{\Delta n(\theta=0)}{n_o} \cos^2(\theta(x, y)) \right]^{\frac{1}{2}} - n_o \right\} dy \\ &\approx \frac{2\pi}{\lambda} \int_{y=0}^l \Delta n(\theta=0) \cos^2(\theta(x, y)) dy = \frac{\pi}{\lambda} \int_{y=0}^l n^3 R(E(x, y)) E^2(x, y) \cos^2(\theta(x, y)) dy \end{aligned} \quad (10)$$

where we define the birefringence as quadratic, but with the electro-optic coefficient also being a function of the electric field. This can be more simply stated as

$$\Phi(x) = \int_{y=0}^l \phi(E(x, y)) \cos^2(\theta(x, y)) dy \quad (11)$$

where $\phi(E(x, y))$ is the phase function from the experimental curve of Figure 2b. Applying the electric field results from FEA we find

$$\Phi(x) \approx \sum_{i=1}^N \phi(E(x, y_i)) \cos^2(\theta(x, y_i)) l_i \quad (12)$$

where N is the number of finite elements that the light ray passes through and l_i is the height of each element.

Mentat FEA software calculates the x and y components of the electric field (in this case we are using a two dimensional model) for four integration points for each element. Taking the average over each element we use the magnitude of the field (i.e. $E = \sqrt{E_x^2 + E_y^2}$), the phase function and the orientation of the index ellipsoid from Equation 4 to get the relative phase change for each element. Integrating the change in phase passing through a column of elements, expressed by Equation 12, we find the change in phase for a light ray passing through a line of elements (see the dotted line in Figure 3). Looking at the series of columns across the electrode gap gives us a phase profile for a plane wave passing through the device.

4. MODELING VS. EXPERIMENTAL PERFORMANCES

Using the modeling procedures discussed in section 3 we determine the calculated phase distribution for a simulated ISE device and compare it to that found experimentally using a fabricated device. The FEA modeling and the experimentally measured results are found to be in good agreement for voltages of less than $V_{2\pi}$ and between electrode gaps of about 500 μm (see Figure 4a). For narrow gaps and higher voltages (see Figure 4b) there is a difference between the modeling and experimental results. One possible explanation would be that the electric field

strength is weaker than the value calculated by the FEA model due to screening effects. These screening effects occur when free carriers (photo-induced or due to surface states) create a space-charge distribution near the electrodes. Our future work will entail investigating these phenomena.

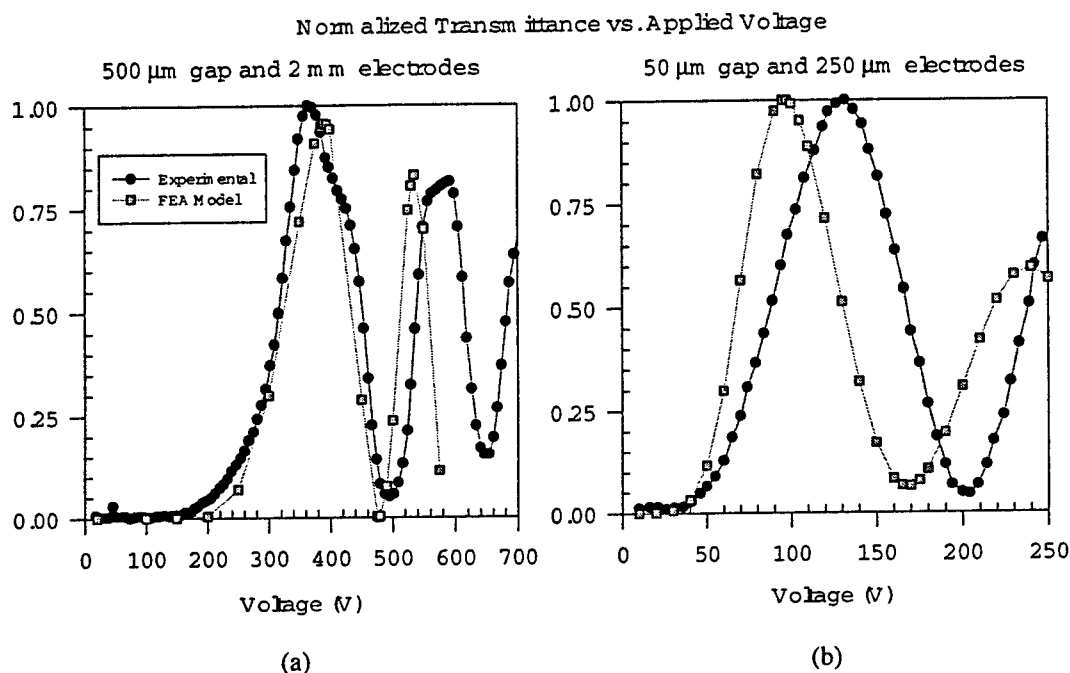


Figure 4. Comparison of FEA model to experimental data for transmission through crossed polarizers. (a) Shows a relatively good fit for V_{π} and $V_{2\pi}$ (i.e. the first maxima and minima) for an electrode gap of 500 μm . Whereas (b) shows a poor fit with a narrow electrode gap of 50 μm .

Currently, in our model, we are compensating for this 'weakening' effect by introducing a constant factor. Consequently, we are able to accurately model the behavior of fabricated devices operated at voltages less than $V_{2\pi}$ (see Figure 5a). Notice that the 'integrated phase' depends linearly on the applied voltage in contrast to the quadratic behavior predicted by previous models.

For surface electrode based devices, the two parameters that are most important are the electrode width and the size of the gap between electrodes. By varying the gap size, and holding the width (160 μm) and voltage (150 V) constant, in our FEA model we have determined that as the gap decreases the phase shift increases proportionately. This is expected based upon the linear relationship of phase to electric field, since the electric field strength scales proportionately with gap size. However, when the gap size becomes smaller than 40 μm there is very little increase in phase modulation. This is due to electric field values being above the 'saturation' level. We also find that as the electrodes increase in size there is a corresponding increase in phase shift, but electrodes wider than 160 μm were also of little benefit (see Figure 5b). The conclusion is that the optimal configuration for this material is a 40 μm gap between

160 μm wide electrodes. It was experimentally determined⁸ that for interdigital surface electrodes on PLZT 8.8/65/35 the optimum configuration is 160 μm electrode widths with 40 μm between electrode gaps. From this excellent agreement we conclude that optimization of design parameters can be done successfully using our FEA model.

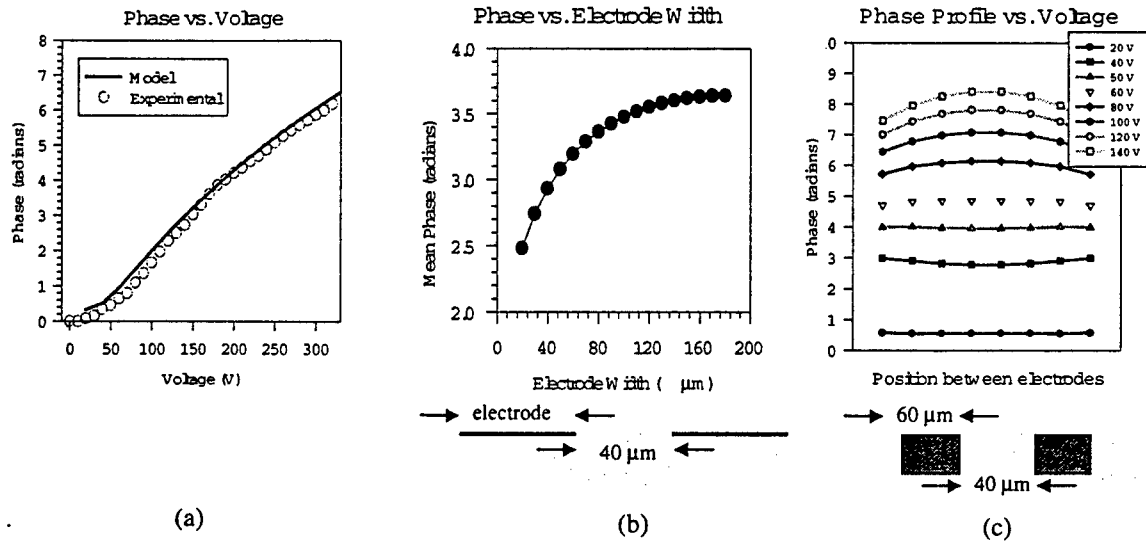


Figure 5. (a) Comparison of the FEA model with experimental data for phase as a function of applied voltage for surface electrodes 160 μm wide with a 40 μm gap (after including a compensating constant factor). (b) Holding the gap width (40 μm) and voltage (150 V) constant, the results of the FEA simulation show that as electrode width approaches 160 μm the increase in phase modulation slows down. (c) The gradient of the phase front for various applied voltages for etched electrodes 40 μm deep. A relatively flat phase profile across the entire gap can be realized around 50 Volts.

Gap Width (μm)	Electrode Width (μm)	Electrode Depth (μm)	V_{π} (V)
80	20	100	56.5
60	15	75	46.0
40	10	50	36.0
20	5	25	30.0

Table 1. The reduction in the necessary applied voltage to achieve π phase modulation highlights the advantages of 'scaling' down the electrode size and spacing.

With our general modeling approach, by varying other parameters within our FEA model we can design and evaluate many device configurations without having to fabricate many devices. In the following we will briefly discuss the optimization of two independent design criteria, embedded electrode geometry and phase uniformity. One of the limitations in using ISE devices is their low transmittance due to a small gap to electrode ratio, i.e. small fill factor. In contrast to surface electrodes, much larger fill factors can be realized by using electrodes that are embedded into the surface. To find the effect of using different etch depths we hold other parameters constant and change the scale of the electrode structure. Using a fill factor of 80% and assuming that a 5:1 aspect ratio is possible in etching PLZT substrate, we observe that V_{π} decreases steadily as the

gap gets smaller (see Table 1), indicating that the decrease in switching energy is proportional to the decrease in gap size. Under the constraint of an 80% fill factor we observe that the linear relationship no longer holds for a gap size below 40 μm . For small geometries the electric fields need to be high due to the short active modulation path length. Consequently, we once again observe the effects of 'saturation' of the phase modulation.

Another important characteristic of an electro-optic modulator is the homogeneity of the modulated phase front. Our FEA modeling allows us to calculate the phase distribution across the aperture between the electrodes. A plane wave passing through the gap between electrodes will attain a phase curvature across the aperture depending on the field distribution. According to our model, for a 50 μm wide gap at 150 Volts, there is a phase difference of approximately 0.7 radians from center to edge. By simulating various electrode geometries and applied voltages we find that by using electrodes 60 μm wide, spaced 40 μm apart and etched 40 μm deep the wave front has an almost perfectly flat π phase profile at 53.6 Volts (see Figure 5c). It is important to note here that we have been analyzing one characteristic at a time. In order to find an optimum device configuration, many coupled characteristics must be taken into account and weighted according to specific device requirements.

5. CONCLUSION

We have used a uniform applied electric field within PLZT in order to experimentally characterize the electro-optic response of the material. This characterization has highlighted the fact that scattering and depolarization effects need to be considered in determining the phase function. Furthermore, electric field distributions obtained using various electrode configurations have been calculated using FEA. These resultant electric fields were integrated with the phase function of the material to determine the characteristic phase modulation of an ISE electro-optic device. The calculated strength of the electric fields has shown us that 'saturation' of phase modulation needs to be considered in device design. We have also found that an electric field 'weakening' effect needs to be factored into the model and in the future we will investigate the cause of this phenomena. After compensating for these various effects we are able to model the behavior of a device as a function of a variety of parameters. We have shown that this model is useful in optimizing individually such device characteristics as increased transmittance and homogeneity of the phase front. Multiple characteristics of such devices that are mutually coupled can also be optimized.

6. ACKNOWLEDGMENTS

We wish to thank the invaluable help of Rebecca Bussjager at Rome Labs, Eddie Rezler from Marc Analysis as well as Rong-Chung Tyan and James Thomas at UCSD. This work is supported by Air Force Rome Laboratory.

7. REFERENCES

1. G.H. Haertling and C.E. Land, 'Hot-Pressed (Pb,La)(Zr,Ti)O₃ Ferroelectric Ceramics for Electrooptic Applications', J. Am. Ceram. Soc., Vol. 54, No. 1, pp. 1-10, 1971.

2. R. Viennet, 'Driving Voltage Calculation for a Ferroelectric Display Device', *J. de Math. et de Phys. Appliquées*, Vol. 29, pp. 715-22, 1978.
3. E.E. Klotin'sh, et al., 'Géometrical Optics of an Electrically Controlled Phase Plate made of PLZT-10 Ferroelectric Ceramic', *Avtometriya*, No. 6, pp. 68-72, 1984.
4. K. Tanaka, et al., 'Analyses of PLZT Electrooptic Shutter and Shutter Array', *Jap. J. Appl. Phys.*, Vol. 24, No. 2, pp. 177-82, February 1985.
5. M. Title and S.H. Lee, 'Modeling and characterization of embedded electrode performance in transverse electrooptic modulators', *Appl. Opt.*, Vol. 29, No. 1, pp. 85-98, January 1990.
6. Q. Wang Song, P.J. Talbot and J.H. Maurice, 'PLZT based high-efficiency electro-optic grating for optical switching', *J. Modern Optics*, Vol. 41, No. 4, pp. 717-27, 1994.
7. B. Mansoorian, *Design and Characterization of Flip-Chip Bonded Si/PLZT Smart Pixels (Ph.D. Thesis)*, Ch. 2, University of California, San Diego, 1994.
8. J. Thomas and Y. Fainman, 'Programmable diffractive optical elements using a multichannel lanthanum-modified lead zirconate titanate phase modulator', *Opt. Lett.*, Vol. 20, No. 13, pp. 1510-12, July 1995.
9. A. Yariv and P. Yeh, *Optical Waves in Crystals*, Ch. 5, John Wiley and Sons, New York, 1984.
10. M. Ivey, 'Birefringent Light Scattering in PLZT Ceramics', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 38, No. 6, pp. 579-84, IEEE, New York, 1991.
11. H. Elgan, 'Excitation of Elastic Surface Waves by Spatial Harmonics of Interdigital Transducers', *IEEE Transactions on Electron Devices*, Vol. 16, No. 16, pp. 1014-17, IEEE, New York, 1969.
12. A. Y. Wu, T. C. Chen and H.Y. Chen, 'Model of electrooptic effects by green's function and summary representation: applications to bulk and thin film PLZT displays and spatial light modulators', *Proceedings of Eighth IEEE International Symposium on Applications of Ferroelectrics*, Eds. M. Liu, A. Safari, A. Kingon and G. Haertling, pp. 600-3, IEEE, New York, 1992.
13. G. Haertling, 'Electro-optic ceramics and devices', *Electronic Ceramics*, Ed. L. Levinson, pp. 371-492, M. Dekker, New York, 1988.

A. 11. Modeling Electric Field Induced Effects in PLZT EO Devices

P. Shames, P.C. Sun and Y. Fainman

TOPS, Spatial Light Modulators, Vol. 14, Eds. Burdge and S.C. Esener, (OSA, Washington, D.C.) pp. 121-3, 1997

Modeling Electric Field Induced Effects in PLZT EO Devices

Paul E. Shames, Pang-Chen Sun and Yeshayahu Fainman

Department of Electrical and Computer Engineering
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0407

Abstract

PLZT 8.8-9.5/65/35 electrooptic (EO) devices subject to electric fields on the order of $1 \text{ V}/\mu\text{m}$ can become highly scattering and depolarizing as well as exhibit high-order EO effects. We describe a method of modeling EO modulators that encounters for these effects. Utilizing these characteristics in simulating surface electrode devices we compare our model to measurements of a fabricated device and find excellent correlation.

Key Words

PLZT modeling, Scattering and Depolarization, Optical Characterization, Mueller Matrices.

Introduction

Applying the advances in electronic and optical computer aided design (CAD) to optoelectronic systems requires precision simulation of electrooptic (EO) devices. In this paper we present a simple but accurate method of modeling EO devices using, as an example, lanthanum-modified lead titanate zirconate (PLZT) with compositions 9.X/65/35. These ferroelectric ceramics have strong EO effects and are cost effective and are therefore excellent candidates

for use in optoelectronic systems. In the absence of field the ceramic is optically isotropic whereas an applied field induces anisotropy and optical birefringence. This electrooptic response has been modeled as a classic Kerr quadratic effect [1] as well as a combination of linear and quadratic effects [2]. However, these techniques fall short of accurately modeling fabricated devices, which will be required for optoelectronic CAD tools.

We have developed a simple method of characterizing a sample of EO material (e.g. PLZT) using a set of equations that relate scattering, depolarization and electrooptic effect to applied electric field. By integrating these responses with the calculation of the field, for given device geometries and driving voltages, we have achieved excellent correlation to experimental data. In this paper we describe our EO characterization and device modeling, with an emphasis on comparisons of computer simulation to fabricated devices.

Transmittance Model and Characterization

Previous work [3] and experimental evidence shows that light passing through PLZT-based devices experiences depolarization due to multiple scattering effects. Therefore, we use the coherency matrix formalism, which describes partially polarized light by taking into account polarized and unpolarized

components, to generate an equation describing the intensity of a monochromatic plane wave transmitted through our apparatus. For example, for linearly polarized light passing through crossed polarizers set at 45° the equation relating transmitted intensity to change in relative phase is given by

$$I = \frac{1}{4} \{A + B + 2C \pm 2\sqrt{AB} \cos(\phi)\} \quad (1)$$

where A and B represent the polarized light intensity measured along the two orthogonal basis and C is the unpolarized component, ϕ is the change in relative phase between the two orthogonal components. Choosing four pairs of angles, for the incident polarizer and analyzer, gives four equations which are used to solve for the relative change in phase ϕ .

By sandwiching our PLZT sample between two large, parallel, gold-plated copper plates we generate a constant field inside and outside the ceramic. The distance between plates and the thickness of the PLZT give us the electric field strength and interaction length, respectively. We experimentally measure the normalized optical intensity for various orientations of the polarizer/analyzer pairs (see Figure 1). For incident light polarized at 45° with a parallel analyzer we see a sinusoidal variation of the intensity as the relative phase is changed with applied voltage, as predicted in Eq. 1. In the measurement of the vertical (0°) and horizontal (90°) components we see the effects of scattering and depolarization as the field is increased.

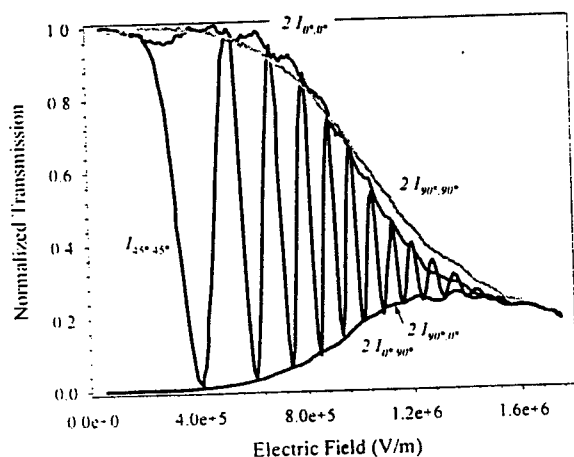


Figure 1. The sinusoidal curve is the transmission through parallel polarizers oriented at 45° to the direction

of the electric field. The mismatch between this curve and the 'envelope' is due to a wedge shaped PLZT sample that results in slightly varying path length within the sampled region.

Using the transmission data from Figure 1, we solve for the change in relative phase ϕ as a function of electric field (see Figure 2). To accurately curve fit the EO data we need to use at least a fifth order polynomial, indicating that the EO behavior of these PLZT ceramics cannot be fully described by the quadratic EO effect.

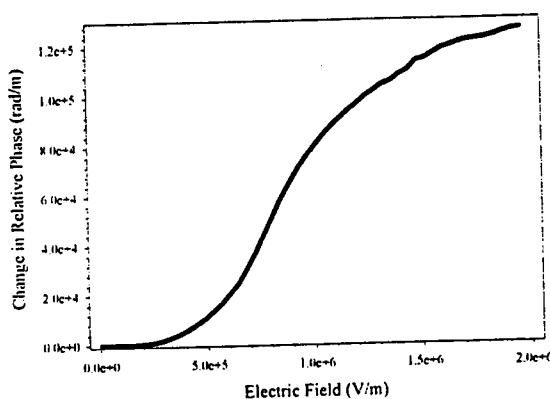


Figure 2. The change in relative phase is shown for the PLZT 8.9/65/35 composition.

Device Modeling

For a plane wave propagating through a PLZT based device we can use a standard index ellipsoid approximation [4] to find the index change for any two orthogonal polarization components of the light. We use a commercial finite element analysis (FEA) tool to calculate the electric field distribution for arbitrary device geometries. By integrating the change in index of refraction, scattering and depolarization data with the calculated electric field components we can determine the change in index as a function of position for any given device geometry.

For incident beam linearly polarized at 45° with respect to the electrodes and using an analyzer placed at -45° we compare our simulation with a fabricated surface electrode device. For this comparison, we used 1 mm CrAu electrodes spaced 500 μm apart and averaged the transmitted intensity from the center 100 μm of the gap. As can be seen in Figure 3, we

get an excellent match of device behavior, including the 'flat' response at low voltage, the reduction of maximum intensity due to scattering and the decrease in contrast due to an unpolarized bias component.

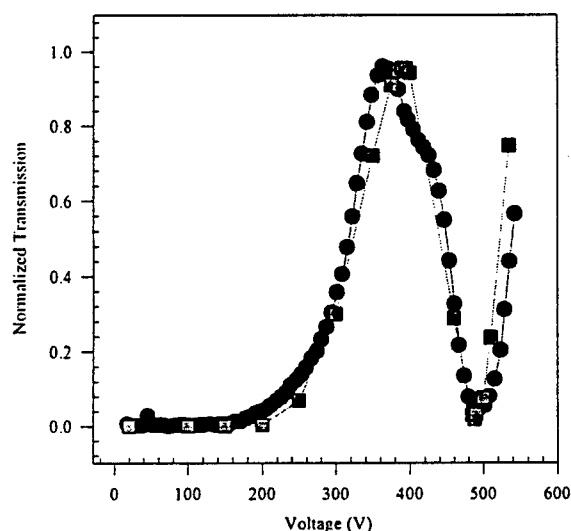


Figure 3. Comparison of measured performance of fabricated surface electrode PLZT device (dots) and computer simulation using new modeling method (line). Device has 2mm wide CrAu electrodes separated by a 500 μm gap and evaporated onto a 380 μm thick wafer.

Conclusion

By using four polarizer/analyzer orientations and taking measurements of optical response versus applied voltage we generate sufficient information to calculate the relative change in index of refraction for PLZT. From these measurements we extract information on the scattering, depolarization and change in relative phase. By incorporating that same information into simulating device behavior we then

accurately model arbitrary geometries for polarization rotation devices. We are currently using this technique to optimize device design for a variety of performance characteristics including maximum contrast, efficiency and minimized cross talk. This empirically-based modeling of electrooptic devices is essential for developing accurate and reliable CAD tools necessary for design of future optoelectronic systems

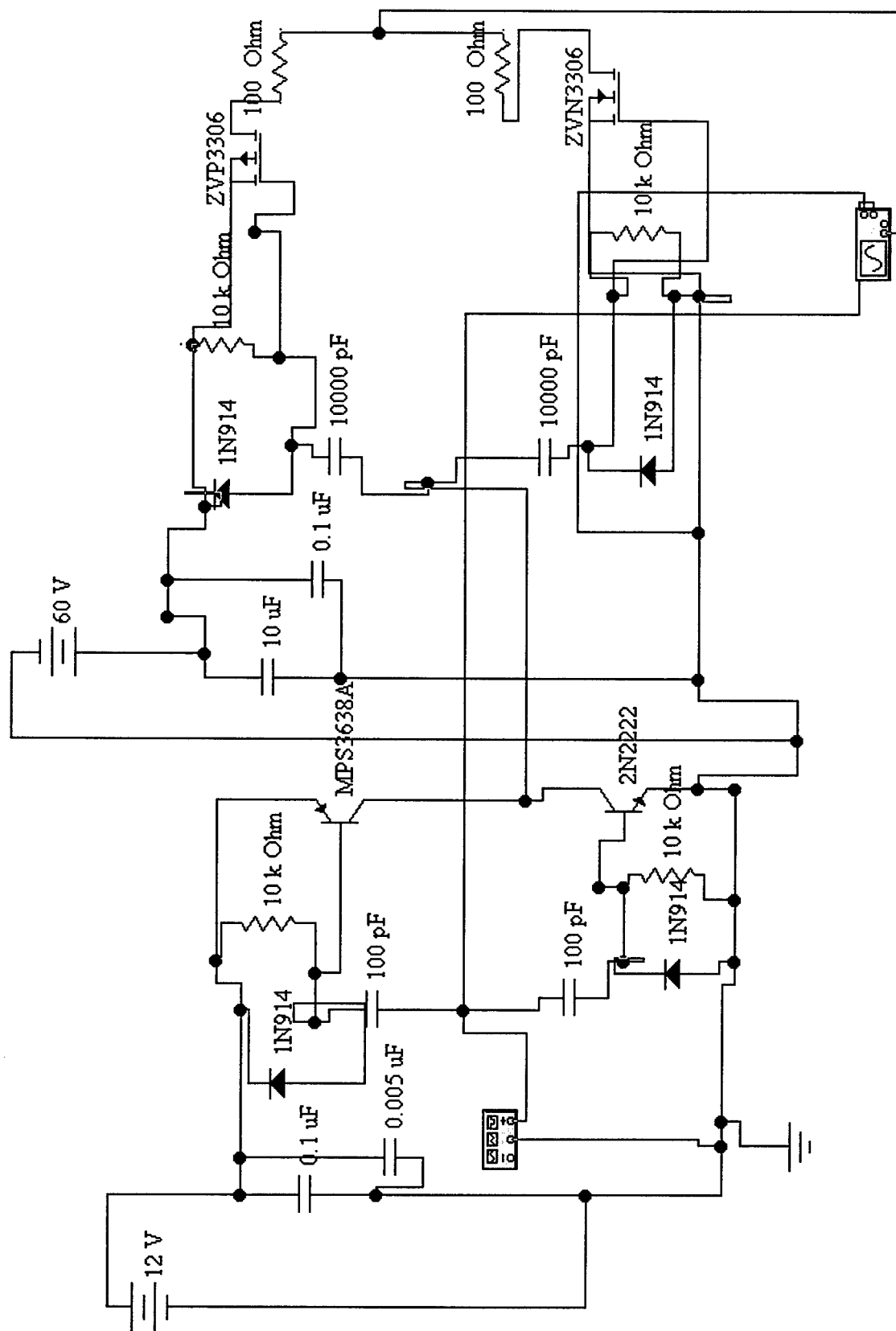
Acknowledgments

We wish to thank Eddie Rezler of Marc Analysis for his assistance with computer simulations. This research is funded in part by AFOSR, NSF and Rome Lab.

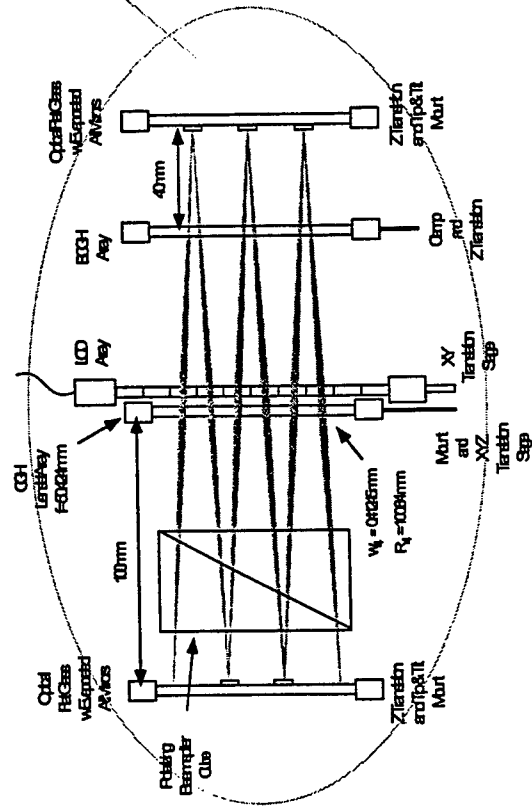
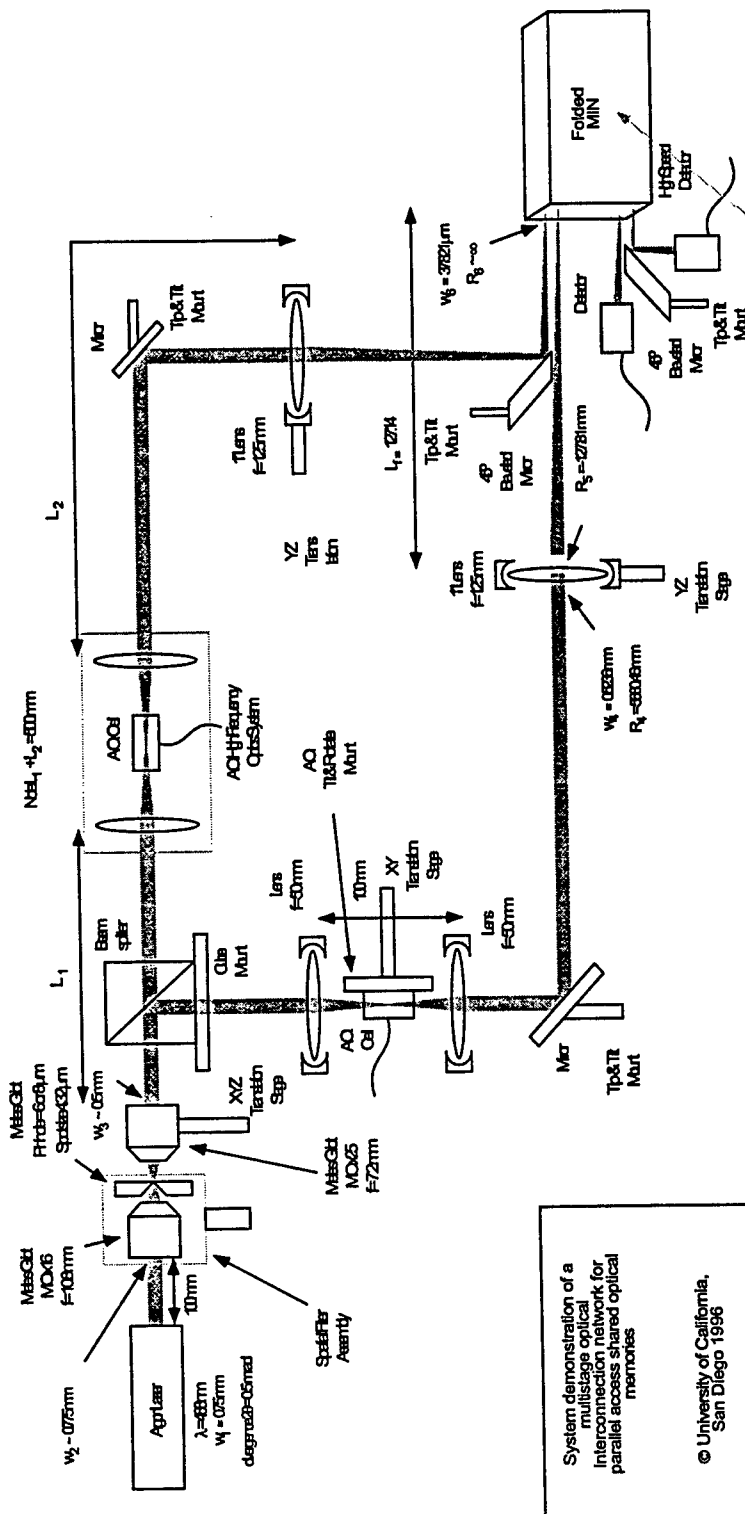
References

1. G. Haertling and C.E. Land, 'Hot-Pressed (Pb,La)(Zr,Ti)O₃ Ferroelectric Ceramics for Electrooptic Applications', *J. Am. Ceram. Soc.*, Vol. 54, No. 1, pp. 1-10, 1971.
2. B. Mansoorian, *Design and Characterization of Flip-Chip Bonded Si/PLZT Smart Pixels (Ph.D. Thesis)*, Ch. 2, University of California, San Diego, 1994.
3. M. Ivey, 'Birefringent Light Scattering in PLZT Ceramics', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 38, No. 6, pp. 579-84, IEEE, New York, 1991.
4. M. Title and S.H. Lee, 'Modeling and characterization of embedded electrode performance in transverse electrooptic modulators', *Appl. Opt.*, Vol. 29, No. 1, 85-98, January 1990.
5. G. Haertling, 'Electro-optic ceramics and devices', *Electronic Ceramics*, Ed. L. Levinson, 371-492, Coté and Bouchard, 1988.

A. 12. Schematic for push-pull MOSFET driver



A. 13. System design and component layout for optical MIN demonstration



A. 14. Folded Free-Space Polarization-Controlled Multistage Interconnection Network

D. Marom, P. Shames, F. Xu and Y. Fainman

Applied Optics, Vol. 37, No. 29, pp. 6884-91 (1998).

Folded free-space polarization-controlled multistage interconnection network

Dan M. Marom, Paul E. Shames, Fang Xu, and Yeshaiah Fainman

We present a folded free-space polarization-controlled optical multistage interconnection network (MIN) based on a dilated bypass-exchange switch (DBS) design that uses compact polarization-selective diffractive optical elements (PDOE's). The folded MIN design has several advantages over that of the traditional transparent MIN, including compactness, spatial filtering of unwanted higher-order diffraction terms leading to an improved signal-to-noise ratio (SNR), and ease of alignment. We experimentally characterize a folded 2×2 switch, as well as a 4×4 and an 8×8 folded MIN that we have designed and fabricated. We fabricated an array of off-axis Fresnel lenslet PDOE's with a 30:1 SNR and used it to construct a 2×2 DBS with a measured SNR of 60:1. Using this PDOE array in a 4×4 MIN resulted in an increased SNR of 120:1, highlighting the filtering effect of the folded design. © 1998 Optical Society of America

OCIS codes: 090.1760, 060.4250, 200.4650, 060.1810, 200.2610, 230.5440.

1. Introduction

As the demand for communication and computing services increases, there is a correlated growth in the need to switch among large numbers of input-output ports that carry high-bandwidth signals. Multistage interconnection networks (MIN's) are an attractive switching architecture because of the minimal number of switching elements required.¹ An optical MIN switching system routing high-bandwidth optical signals can play an important role in the development of ultrahigh-bandwidth interfaces with high-capacity parallel-access optical memories as well as for memory distribution. Various optical MIN system implementations have been reported, including guided-wave optics that use LiNbO_3 switches,² free-space optics with optoelectronic switches based on symmetrical self-electro-optic effect devices,³ and transparent switches based on polarization modulators.⁴⁻⁸ Transparent switches, in which an optical signal propagates without regeneration, do not introduce the additional limitations of optoelectronic device cost, speed, and

power. Totally transparent systems depend on a centralized controller's performing the routing algorithm and serving user requests. Recently such a 4×4 free-space optical MIN based on polarization-selective diffractive optical elements (PDOE's) was demonstrated.⁹ Optical MIN's have inherent insertion losses that can limit system scalability. However, attenuation can be compensated for by use of optical fiber amplifiers. Polarization-dependent systems can also take advantage of the recent advances in polarization-maintaining fibers or compensation in single-mode fibers.¹⁰ Other performance metrics that can limit MIN scalability are optical cross talk, system compactness, system stability, and ease of alignment.

In this paper we describe the design and implementation of a free-space optical MIN by use of a novel folded dilated bypass-exchange switch (DBS) built of PDOE's that addresses these limitations. The use of a DBS allows for the elimination of the first-order cross talk that results from inaccuracies of polarization rotation¹¹ and diffractive optical element fabrication errors. By utilizing the three-dimensional functionality of the optical elements, one can then stack the DBS elements in the vertical dimension by folding (by use of a mirror plane) the switch along a central line of symmetry. The interconnection among multiple DBS's can also be folded, forming an optical MIN. This results in a resonator-type structure in which all the switching elements are distributed on a plane, providing a highly compact optical system that can easily be

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0407.

Received 9 January 1998; revised manuscript received 1 June 1998.

0003-6935/98/296884-08\$15.00/0

© 1998 Optical Society of America

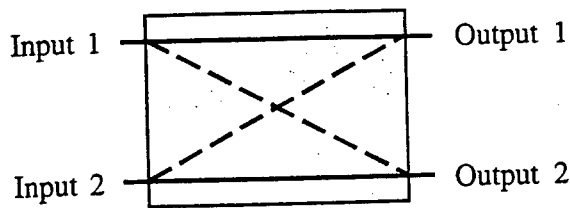


Fig. 1. BES functionality block diagram: solid lines, bypass mode in which the input to channel 1 goes to the output of channel 1; dashed lines, exchange mode in which the input to channel 1 goes to that output of channel 2 and vice versa.

aligned. Additionally, the use of space-variant diffractive optics permits the implementation of any interconnection topology corresponding to arbitrary network architectures. Finally, the use of micromirrors to fold the DBS allows for spatial filtering of the undesired diffraction orders, thereby decreasing the cross talk.

In Section 2 we review MIN switching concepts based on PDOE's. In Section 3 we introduce the folded implementation of the DBS by use of PDOE's. In Section 4 we discuss system design and component fabrication and characterization. Performance evaluation of a 2×2 folded DBS as well as multistage 4×4 and 8×8 folded optical MIN's is presented in Section 5. Finally, in Section 6 we summarize our research and discuss conclusions.

2. Free-Space Optical Multistage Interconnection Networks with Bypass-Exchange Switches

The basic structure of a MIN has alternating arrays of fixed interconnection patterns and switching modules, typically bypass-exchange switches (BES's). The MIN architecture determines the fixed interconnection pattern between switching stages and the number of stages implemented. Various MIN architectures are differentiated by the number of switching stages, complexity of routing algorithms, and network protocols. An optical MIN implemented with space-variant lenslets in free space permits the implementation of arbitrary network architectures and interconnection patterns. Here we demonstrate a folded 8×8 MIN based on a banyan architecture. However, the cross-talk and fabrication issues addressed also apply to any other architecture implementation.

The BES is a 2×2 switch with two allowed states: bypass, in which the signals of the two channels are unchanged (i.e., the input to channel 1 goes to the output of channel 1), and exchange, in which the signals go to the opposite output ports or channels (Fig. 1). Other possible states, known as broadcast and combine, are not considered in this application. A possible optical implementation of the BES uses the polarization state for switching. Two orthogonally polarized light beams are controlled by a polarization rotator to set the state of the switch and the polarization-selective optical elements (e.g., polarization beam-splitter cubes, birefringent crystals, and PDOE's) to direct the beams.

A birefringent computer-generated hologram¹²

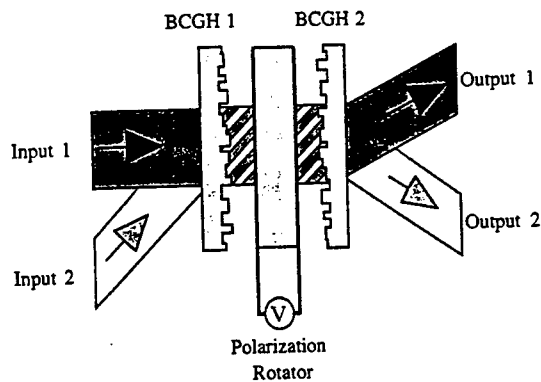


Fig. 2. Optical implementation of the BES: the first BCGH collimates the two input beams and the second BCGH directs the output beams depending on the polarization states. The voltage on the polarization rotator determines the state of the switch.

(BCGH) is an example of a compact and efficient PDOE. A BCGH element has an independent impulse response for each state of the two orthogonal linear polarizations, achieved by the etching of phase encodings into birefringent media. A compact 2×2 optical BES that uses BCGH elements has been demonstrated⁸ (Fig. 2): The first BCGH element combines and focuses two inputs into the polarization rotator, which either exchanges their polarizations or does not. The second BCGH separates and directs the outputs to different destinations according to their polarization states.

Inaccuracies of the polarization rotator and the BCGH fabrication can result in cross talk in this implementation of the BES. The polarization rotator can be characterized by an associated error of δ in the rotation angle, which results in a cross-talk term proportional to $\sin(|\delta|)$. The BCGH elements can be described by an associated cross talk ϵ that is due to fabrication errors such as etch depth and misalignment among multiple masks. The combined cross-talk component at the output of the BES is proportional to $|\delta| + |\epsilon|$, assuming that $\delta, \epsilon \ll 1$. The signal-to-noise ratio (SNR) of a MIN can be described by

$$\text{SNR} = \log_{10} \left(\frac{1}{\delta_c} \right) - \log_{10} S, \quad (1)$$

where $\delta_c = |\delta| + |\epsilon|$ and S is the number of interconnection stages.⁹ For increased scalability of the MIN network size (i.e., S is growing) the cross talk δ_c of each stage must be reduced, yielding the SNR necessary to support the desired bit-error rate.¹³

An improvement in cross-talk performance can be achieved by use of a DBS.¹¹ The DBS, which has two input and two output signals, comprises four 1×2 switches coupled together. The structure of the DBS guarantees that each 1×2 switch has only one signal propagating through it and that the majority of the cross-talk terms exit from the unutilized output ports. It can be shown that the remaining cross talk δ_c is now reduced to $\delta^2 + \epsilon^2$. Under the assumption

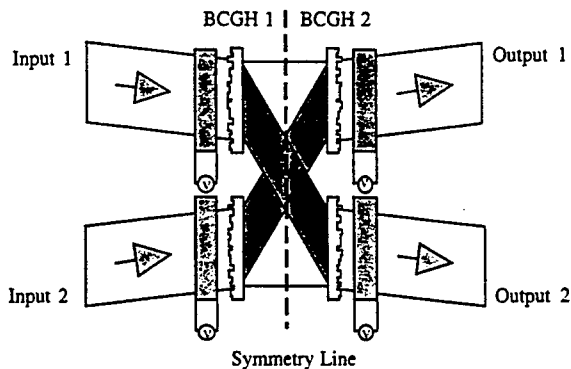


Fig. 3. Optical implementation of the DBS. Each BCGH element performs 1×2 switching, depending on the state of the polarization-rotator element. The input and the output states of the two channels are identical, permitting filtering with a polarizer of linear cross talk at the output.

that $\delta, \epsilon \ll 1$, this performance is a significant improvement over that of a conventional BES.

A free-space DBS can be implemented with a combination of four BCGH's and four polarization-rotator elements (Fig. 3). Depending on the state of the first set (i.e., the first elements in channels 1 and 2) of polarization rotators, the first set of BCGH elements defines the bypass or exchange functionality of the switch. The second set of BCGH lenslets directs the output beams to the next DBS array (for a multistage configuration), where the direction is dictated by the interconnection architecture that is being implemented. The second set of polarization rotators returns the output polarizations to their original input states.

Unlike in the BES, in a DBS the polarization state of each channel remains independent of the others. In our case we specify the input and the output beams to have identical polarization states. Therefore, by placing a polarizer at the output of the DBS, we can eliminate the linear cross talk. In this case the four polarization-rotator elements will always be in the same states, i.e., all ON or all OFF.

3. Folded Dilated Bypass-Exchange Switch and Optical Multistage Interconnection Network

The DBS's complexity, although it mitigates linear cross-talk problems, increases the number of components required for it to have the same functionality as the BES. However, one can reduce the complexity of these switches by taking advantage of the symmetry of the DBS (dashed line in Fig. 3) and the three-dimensional functionality of our free-space optical elements. One does this by introducing a propagation-direction component along the vertical axis, i.e., a small incidence angle, as well as by placing a mirror at the line of symmetry (Fig. 4). The input beams will pass through a rotator-BCGH combination at one elevation and react according to the encoded information at that location, switching information in the horizontal direction. On reflection

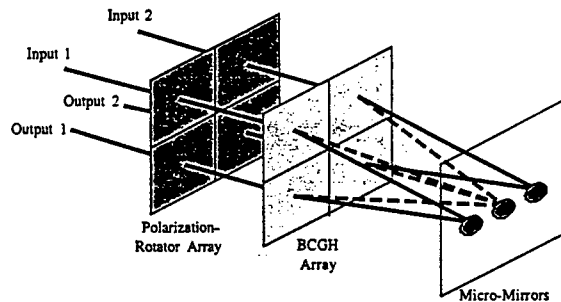


Fig. 4. Folded optical DBS. Similar elements (i.e., BCGH's, polarization rotators) are placed in two-dimensional arrays. Micromirrors reflect only the desired diffraction order and filter out the unwanted orders. The four polarization rotators are always in the same state and can be replaced by one larger-sized element.

from the mirror the beam passes through another BCGH-rotator combination but at a lower elevation (Fig. 4). By folding the switch in this manner, one locates similar elements (e.g., BCGH lenslets) in the same plane. Therefore a single DBS can be fabricated by use of a mirror and 2×2 arrays of BCGH's and polarization-rotator elements. The four polarization rotators, which are always in identical states, can be replaced with one larger polarization-rotator element. However, if we wish to consider other switching functionalities such as broadcast and combine states, the four polarization rotators have to be controlled separately.

The advantage of this folding technique is further enhanced when it is applied to an optical MIN. When a mirror is placed at the output of the first folded DBS, the beam will reflect back at a lower elevation and be coupled into subsequent DBS's located below the first. In this manner all similar elements of multiple DBS's can be combined into two-dimensional arrays, minimizing the number of components required for the entire MIN: a single BCGH array, a polarization-rotator array, and a pair of folding micromirror arrays. A folded optical MIN is packaged as a resonator in which each round trip represents a stage and all stages are stacked vertically (Fig. 5). An input signal beam enters the system at a small angle and reflects through a prescribed number of stages before exiting in the desired spatial output channel.

Implementation of a free-space optical MIN by use of this folding technique and BCGH space-variant lenslets presents several unique advantages:

(a) Arbitrary architecture: The use of space-variant lenslets in each polarization-selective element allows for the design of arbitrary connection patterns such that any multistage network topology can be implemented. In the folded optical MIN the number of channels and the interconnection architecture used dictate the size of the arrays but do not increase the number of components. For example, an 8×8 optical MIN architecture (of $\log_2 8 = 3$

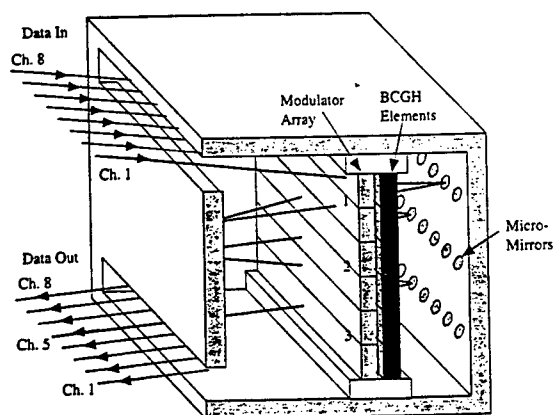


Fig. 5. Folded 8×8 optical MIN with one input beam, shown propagating from channel 1 to channel 5. In this example there are three stages of DBS's, which require three round-trip travels in the micromirror cavity.

stages) requires arrays of size 8×6 in BCGH's and polarization-rotator elements.

(b) Spatial filtering: A further advantage of the folded switch is that the mirror planes can also implement filtering functionality to increase SNR performance further. BCGH's are diffractive elements that, depending on the element design, can produce undesired diffraction terms. When continuous mirrors are used these undesired orders can propagate within the MIN, resulting in additional cross-talk noise at the output. However, if micromirrors deposited upon a transparent substrate are used, only the desired diffraction orders from the BCGH's will

reflect back for further propagation, while the unwanted noise terms exit the system.

(c) Alignment: The arrangement of the optical elements in a two-dimensional array format also allows for relatively simple alignment of the system components. Correct alignment will dictate that the beams land on the correct elements during each pass through the cavity. The displacement of each beam from its correct position and the size of the beam at the BCGH elements (i.e., larger or smaller than the predicted size at the element) will indicate which optical elements (BCGH's, micromirrors, etc.) are incorrectly positioned. Inasmuch as the micromirror planes are mostly transparent, beam propagation within the cavity can be viewed with external imaging optics and a CCD camera. The beam size and position can therefore be monitored *in situ*, allowing for accurate alignment of optical elements and mirror planes.

4. System Design, Component Fabrication, and Their Characterization

To demonstrate a folded system, we designed and constructed an 8×8 optical MIN system (Fig. 6) based on a fully connected banyan architecture.⁹ The design process of the system incorporates the following criteria: (i) maximization of the number of rings in off-axis Fresnel lenslets, (ii) minimum feature size of diffractive elements determined by the available fabrication technologies, and (iii) separation of diffractive-order beams at the micromirror plane. We developed a system-modeling tool by using Gaussian beam analysis of the stable mode of the micromirror-based (Fig. 5) cavity that calculates

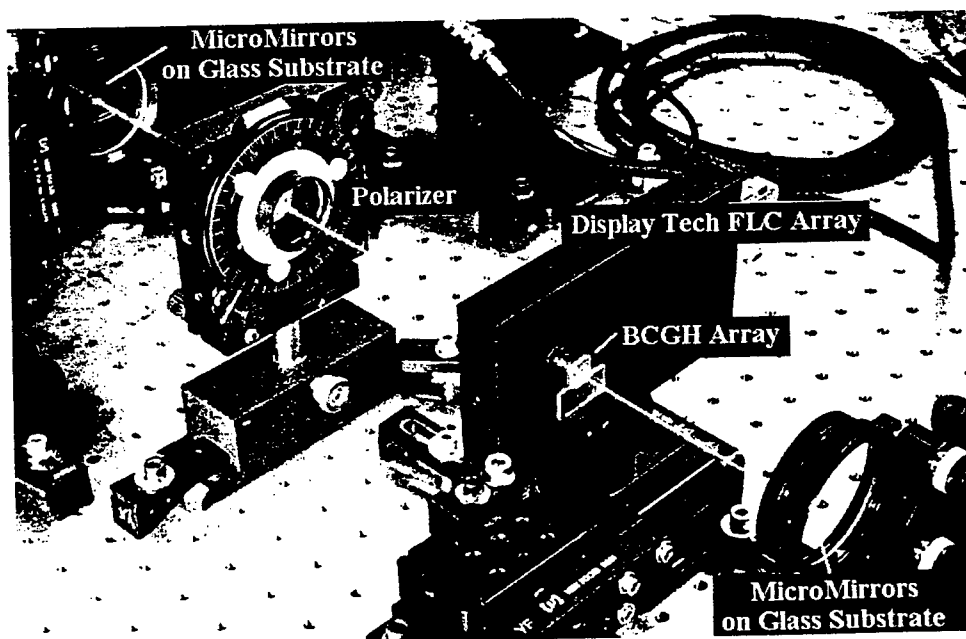


Fig. 6. Photograph of the experimental demonstration system of the folded optical MIN. The cavity is defined by two micromirror arrays deposited upon glass substrates. The polarizer is used to filter out the undesirable linear cross talk that is due to incorrect polarization rotation.

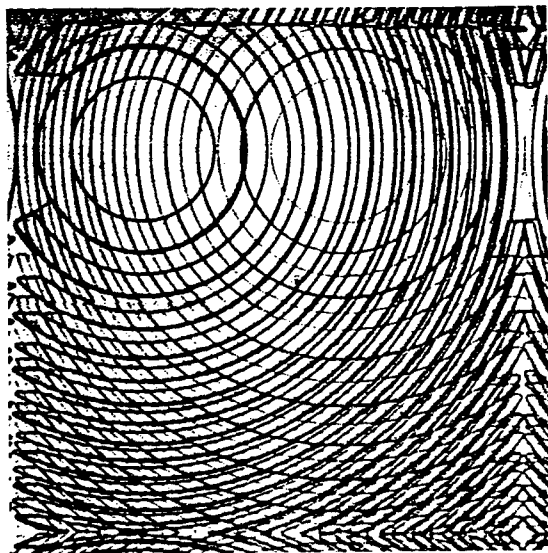


Fig. 7. CCD Image of a single BCGH element within an 8×6 array, showing multifunctional superposition of polarization-selective Fresnel lenses.

these three parameters for a given cavity dimension. We found the optimal size of the cavity by varying the cavity dimensions over our design space and maximizing the above criteria.

The Gaussian beam spot size¹⁴ at the BCGH lenslet plane was used as a limiting design constraint because at that location the beam size is largest. For a spot size greater than the lenslet (i.e., array pitch), optical power would leak into adjacent elements and give rise to cross talk. A spot size much smaller than the BCGH lenslet would result in a diminished diffraction efficiency. Our pitch size of 1 mm was determined by the dimensions of the pixel size of the polarization-rotator array used in our experiments. A beam spot size of 0.825 mm was used and provided minimal cross talk, high diffraction efficiency, and

high power throughput (97% of the beam energy is contained in the $1 \text{ mm} \times 1 \text{ mm}$ square). Accounting for beam propagation through multiple optical elements (BCGH, polarization rotator, etc.) yielded a lenslet focal length of 85.1 mm with a $300\text{-}\mu\text{m}$ waist size at one mirror plane and a $100\text{-}\mu\text{m}$ waist size at the other. The cavity length is 407 mm, with the BCGH element placed 107 mm from the back side mirror. The 1-mm pitch of the optical elements also dictates that the input light beam have an incidence angle of 0.4° . The polarization-rotator array is placed adjacent to the BCGH array to best match the 1-mm pitch of the ferroelectric liquid-crystal (FLC) elements.

Each of the designed BCGH lenslets functions as two independent off-axis Fresnel lenses for the two orthogonal polarization states (Fig. 7), whose offset is dictated by the deflection angle required by the interconnection pattern. The largest deflection angle for our 8×8 banyan network is 0.8° , corresponding to shifting the beam by 3 pixels. The BCGH was designed by use of the multiple-order delay approach¹⁵ and fabricated in an YVO_4 crystal selected for its high value of birefringence. The advantages of using off-axis lenses include the following results: (i) The unwanted zero-order diffraction term can be filtered out of the cavity. Because the diffraction into the zero order is more sensitive to fabrication errors, which results in a strong unwanted residual component, we find that maximum extinction ratios can be attained by use of the first-order diffraction terms. (ii) The unwanted higher-order diffraction light is dispersed over a large area of the micromirror plane and is *not* focused onto the micromirrors [Fig. 8(a)]. The amount of optical power incident upon adjacent micromirrors (i.e., noise) and reflected back into the system is determined by the ratio of the area of the micromirror to the area of the diffracted order at the plane of the mirrors [Fig. 8(b)].

The measured first-order diffraction efficiencies of

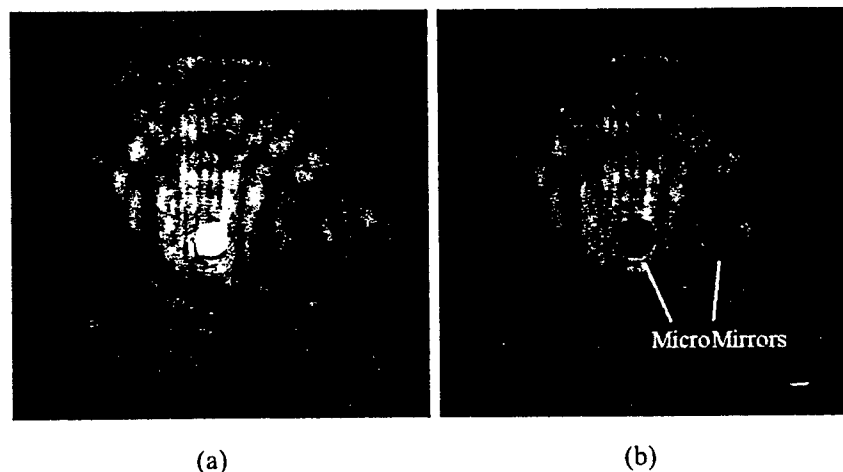


Fig. 8. CCD pictures of (a) the output of one BCGH lenslet element for one polarization state, showing the focused first-order light and the unfocused higher diffraction orders, and (b) micromirrors (the dark circles), reflecting only first-diffraction-order light, permitting higher-diffraction-order light to exit the system.

the binary phase-level BCGH elements are 33% for vertical polarization and 35% for horizontal polarization, with extinction ratios better than 60:1. We also fabricated and tested four-phase-level BCGH diffractive elements, which require greater cumulative etch-depth errors because they have a more complicated fabrication process. These elements yielded 30:1 extinction ratios and diffraction efficiencies of 43% for vertical polarization and 46% for horizontal polarization. The relatively low extinction ratio and diffraction efficiency (compared with a theoretical efficiency of 80.5% for a four-phase-level diffractive optical element) is due primarily to etch-depth inconsistencies in the multistep fabrication process, which are described by the error term ϵ [see Eq. (1)].

The patterned arrays of micromirrors were etched onto a mirror fabricated by the evaporation of aluminum film onto optical flats, for which the average measured reflectance of the mirrors was 92%. The circular micromirrors have diameters of 400 and 150 μm , slightly larger than the calculated beam diameter at the mirror planes.

5. Multistage Interconnection Network System Experimental Characterization

Experimental testing of our 8×8 MIN system was performed with a 488-nm cw Gaussian beam generated by an argon laser. For initial testing we used two optical input channels, one with a dc signal and the other modulated by a NEOS Model N71003 acousto-optic (AO) cell. The polarization state of the beam as it propagates through the network is controlled by the two-dimensional array of FLC polarization rotators (DisplayTech, Model $10 \times 10\text{B}$). Reconfiguration of the FLC elements is under computer control, with a maximum switching speed (i.e., frame rate) of 0.2 ms. The output signals are measured by high-speed silicon p-i-n detectors.

For an 8×8 folded MIN the beam makes three round trips in the cavity (i.e., three layers of BES's). By diverting the beam after one or two passes, we are able to use the same experimental system to test the performance of a 2×2 (single DBS switch) or a 4×4 network. Using the binary phase-level BCGH elements, we measured the performance of a single DBS switch, which yielded extinction ratios of greater than 250:1. The extinction ratio for the DBS is significantly better (4:1) than those of the individual BCGH elements, which shows how the DBS eliminates significant cross talk. However, because of the low diffraction efficiency of the binary phase elements, the DBS switch has an insertion loss of approximately -11 dB and was not suitable for the multistage system experiments.

Figure 9 shows the output from a single DBS, constructed with the higher-efficiency four-phase-level BCGH, as it reconfigures between the bypass and the exchange modes at a 2-kHz rate (i.e., 500- μs packets). The AO signal is modulating one of the input signals with a square wave at 40 kHz (we used this relatively slow input signal to permit simultaneous oscilloscope visualization of both AO and FLC reconfiguration

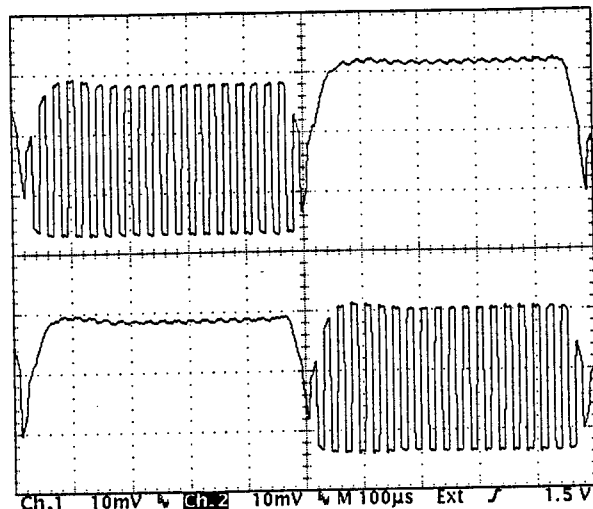


Fig. 9. Output signals for a single folded DBS (2×2 switch) with two input signals: a dc signal and a 20-kHz signal. The switch is reconfiguring at a 1-kHz rate, limited by the 100- μs characteristic rise time of the FLC. The measured average SNR is 57:1.

frequencies). The top and the bottom traces in Fig. 9 show the outputs of channel 1 and channel 2, respectively. Both configurations (i.e., bypass and exchange) produce extinction ratios (defined as the ratio between the ON state and the OFF state when one input signal is present) greater than 59:1 and a SNR (defined as the ratio of the signal to the noise at the same output, i.e., cross talk between two input signals) of greater than 57:1. Results of using signals ranging from dc to 10 MHz show similar SNR's, highlighting the optical transparency of the system. The extinction ratio improvement for the DBS is only 2:1 better than the extinction ratio of the diffraction orders of the four-phase-level elements used. This result is attributed to the much stronger cross talk that is due to fabrication errors (ϵ) seen in these elements. The higher diffraction efficiency of the four-phase-level BCGH reduced the insertion loss of the DBS to approximately -9 dB.

By allowing the beams to propagate two round trips through the cavity (by use of the four-phase-level BCGH elements), we experimentally characterized a multistage 4×4 system. Using a single dc input signal that switched among four output channels, we measured an average SNR of 90:1 and an extinction ratio of 120:1 (a 4:1 improvement over the individual BCGH elements). To investigate cross talk further, we introduced a second input signal. The measured output amplitudes are shown in Fig. 10. Most notable is that the output intensities vary depending on the output as well as the input channel, a result that might occur because of the variation of the lenslet diffraction efficiencies for different polarization states. The minimal average (i.e., the weakest output signal to the strongest output noise) SNR is 87:1.

The complete 8×8 interconnection system was

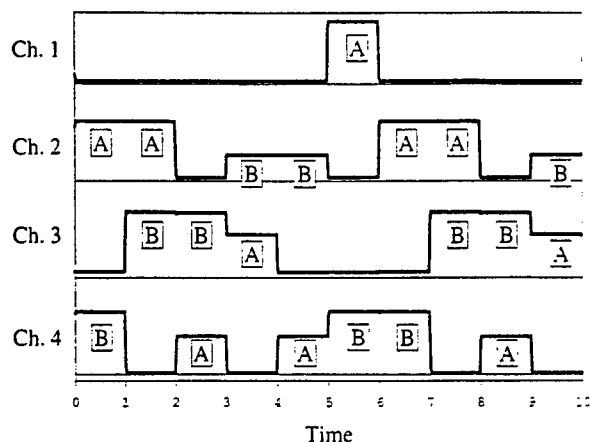


Fig. 10. Output from a 4×4 switch with two input signals (A and B) routed to the four output channels by a host computer controller. The average SNR is 120:1, which is twice as great as that of the single DBS performance, highlighting the filtering capabilities of the folded MIN configuration.

characterized with a single dc input signal that switched among all eight output channels. The output signals were relatively weak and were therefore imaged onto a CCD camera (which integrates over time) for detection (see Fig. 11). The average measured SNR was better than 30:1. This relatively low SNR can be attributed to the strong background noise and the small dynamic range of the CCD device used. We performed similar measurements by using two input signals that switched among all eight output channels that gave similar SNR results.

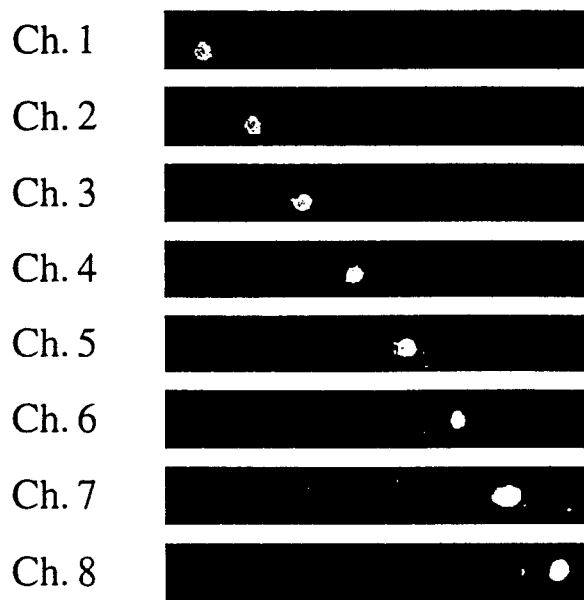


Fig. 11. CCD time-sequenced images showing a single input dc signal routing among eight output channels. The measured average SNR of 30:1 is derived from the CCD pixel values. This relatively low value might be due to the poor dynamic range and the background noise of the CCD device.

6. Discussion

We have described the design, fabrication, and testing of an optical MIN by using a novel folded architecture and compact polarization-selective BCGH elements. The design process determines the optimal system dimensions, which are constrained primarily by the limits of the diffractive optical fabrication facilities available. We have demonstrated how the folded design allows for the elimination of the first-order cross talk, ease of MIN system alignment and packageability, as well as filtering of unwanted high diffraction orders. The use of space-variant lenslets also allows for the implementation of arbitrary MIN architectures. The folded DBS (i.e., 2×2 switches) improved the extinction ratio compared with those of the single BCGH diffractive elements, binary and four-phase level, used. Further improvements in filtering out cross talk were seen in the 4×4 interconnection system, with measured extinction ratios of 120:1 (i.e., when input signals pass through two layers of DBS switches).

Traditional nonfolded MIN systems are planar by design and occupy an area proportional to their number of stages. However, such is not the case for our folded MIN, in which the stages are stacked vertically. The footprint that the system occupies depends on the length of the mirror cavity, which is dictated by the focal length and the deflection angle of the BCGH Fresnel lenslets. For off-axis lenslets the maximum deflection angle will be constrained by the minimum feature size of the fabrication process. However, for feature sizes smaller than five wavelengths,¹⁶ the diffraction efficiency can be adversely affected.

When the N input channels are arranged in a $1 \times N$ vector form, the number of elements in the polarization rotator and BCGH arrays scales as N in the horizontal dimension and as $\log(N)$ in the vertical. As N increases, the ratio of the width to the height of the arrays is increased. For large N a relatively wide system can result, which will require relatively large deflection angles. An alternative strategy for maintaining system compactness is to arrange the N input channels in a rectangular array form (i.e., M rows of length N/M). This procedure would redistribute pixels from the same stage into multiple rows and would result in a more symmetric system that could significantly reduce the maximum degree of the deflection angles required.

The optical transparency of the MIN allows for transmission of very high data-rate signals. We tested our folded MIN with signals from dc to 10 MHz (the limit of our AO cell modulation speed) and found no change in the SNR or the extinction ratio. We expect system performance to be constant for signal bandwidths into the megahertz range. The limiting factor in interconnection reconfiguration is the rise and fall times of the employed FLC polarization-rotation elements. State-of-the-art FLC response times have approximately 10- μ s rise times, but other

electro-optic materials could provide devices with orders-of-magnitude faster response times.¹⁷

The efficiency of the BCGH elements is the limiting factor in increasing the SNR of the multistage system. In our case the poor performance of the Fresnel lenslets can be attributed to two factors: (i) inaccurate etching depths owing to ion-etching device inconsistencies and (ii) use of the multiple-order delay approach to fabrication of BCGH elements, which has increased sensitivity to etch-depth errors. Significantly higher diffraction efficiencies can be expected with improved etching facilities and the use of other BCGH design approaches, such as dual-substrate¹⁸ and form birefringent elements.¹⁹

The use of high-efficiency diffractive elements would also greatly reduce the insertion losses of these systems. For example, using 32-phase-level BCGH Fresnel lenslets with 97% diffraction efficiency²⁰ and dielectric mirrors with 99% reflectance as well as antireflectance coating of all optical surfaces would result in insertion losses less than -1 dB for each stage. For an interconnection system with 10 stages, which allows for 1024 input channels, the total insertion loss would be approximately -7 dB.

This research is funded in part by the National Science Foundation, the U.S. Air Force Office of Scientific Research, and the Rome Laboratory. Dan Marom acknowledges the support of the Fannie and John Hertz Foundation.

References

1. N. K. Ailiwadi, "Photonic switching architectures and their comparison," in *Frontiers of Computing Systems Research*, S. K. Tewksbury, ed. (Plenum, New York, 1990), Vol. 1, pp. 129-186.
2. T. Sawano, S. Suzuki, and M. Fujiwara, "A high-capacity photonic space-division switching system for broadband networks," *J. Lightwave Technol.* **13**, 335-340 (1995).
3. F. B. McCormick, T. J. Cloonan, F. A. P. Tooley, A. L. Lentine, J. M. Sasian, J. L. Brubaker, R. L. Morrison, S. L. Walker, R. J. Crisci, R. A. Novotny, S. J. Hinterlong, H. S. Hinton, and E. Kerbis, "Six-stage digital free-space optical switching network using symmetric self-electro-optic-effect devices," *Appl. Opt.* **32**, 5153-5171 (1993).
4. G. A. DeBiase, "Optical multistage interconnection networks for large-scale multiprocessor systems," *Appl. Opt.* **27**, 2017-2021 (1988).
5. K. M. Johnson, M. R. Surette, and J. Shamir, "Optical interconnection network using polarization-based ferroelectric liquid crystal gates," *Appl. Opt.* **27**, 1727-1733 (1988).
6. K. Noguchi, T. Sakano, and T. Matsumoto, "A rearrangeable multichannel free-space optical switch based on multistage network configuration," *J. Lightwave Technol.* **9**, 1726-1732 (1991).
7. D. M. Marom and D. Mendlovic, "Compact all-optical bypass-exchange switch," *Appl. Opt.* **35**, 248-253 (1996).
8. F. Xu, J. E. Ford, and Y. Fainman, "Polarization-selective computer-generated holograms: design, fabrication and applications," *Appl. Opt.* **34**, 256-266 (1995).
9. A. V. Krishnamoorthy, F. Xu, J. E. Ford, and Y. Fainman, "Polarization-controlled multistage switch based on polarization-selective computer-generated holograms," *Appl. Opt.* **36**, 997-1010 (1997).
10. F. Heismann, A. F. Ambrose, T. O. Murphy, and M. S. Whalen, "Polarization-independent photonic switching system using fast automatic polarization controllers," *IEEE Photon. Technol. Lett.* **5**, 1341-1343 (1993).
11. K. Padmanabhan and A. Netravali, "Dilated networks for photonic switching," *IEEE Trans. Commun.* **COM-35**, 1357-1365 (1987).
12. J. E. Ford, F. Xu, K. Urquhart, and Y. Fainman, "Polarization-selective computer-generated holograms," *Opt. Lett.* **18**, 456-458 (1993).
13. K. S. Urquhart, P. Marchand, Y. Fainman, and S. H. Lee, "Diffractive optics applied to free-space optical interconnects," *Appl. Opt.* **33**, 3670-3682 (1994).
14. A. Yariv, *Quantum Electronics* (Wiley, New York, 1989), Chap. 6.
15. F. Xu, R. C. Tyan, and Y. Fainman, "Single-substrate birefringent computer-generated holograms," *Opt. Lett.* **21**, 516-518 (1996).
16. D. A. Pomet, M. G. Moharam, and E. B. Grann, "Limits of scalar diffraction theory for diffractive phase elements," *J. Opt. Soc. Am. A* **11**, 1827-1834 (1994).
17. J. A. Thomas, M. Lasher, Y. Fainman, and P. Soltan, "A PLZT-based dynamic diffractive optical element for high speed random-access beam steering," in *Optical Scanning Systems: Design and Application*, L. Beiser and S. F. Sagan, eds., *Proc. SPIE* **3131**, 124-132 (1997).
18. N. Nieuborg, A. Kirk, B. Morlion, H. Thienpont, and I. Veretennicoff, "Polarization-selective diffractive optical elements with an index-matching gap material," *Appl. Opt.* **36**, 4681-4685 (1997).
19. F. Xu, R.-C. Tyan, P.-C. Sun, Y. Fainman, C.-C. Cheng, and A. Scherer, "Form birefringent computer-generated holograms," *Opt. Lett.* **21**, 1513-1515 (1996).
20. E. Pawlowski and B. Kuhlow, "Antireflection-coated diffractive optical elements fabricated by thin-film deposition," *Opt. Eng.* **33**, 3537-3546 (1994).

***MISSION
OF
AFRL/INFORMATION DIRECTORATE (IF)***

The advancement and application of information systems science and technology for aerospace command and control and its transition to air, space, and ground systems to meet customer needs in the areas of Global Awareness, Dynamic Planning and Execution, and Global Information Exchange is the focus of this AFRL organization. The directorate's areas of investigation include a broad spectrum of information and fusion, communication, collaborative environment and modeling and simulation, defensive information warfare, and intelligent information systems technologies.